

PI 1057486

REC'D 05 SEP 2003

WIPO

PCT

## THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office

August 28, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM  
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK  
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT  
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A  
FILING DATE.

APPLICATION NUMBER: 60/398,332

FILING DATE: July 24, 2002

RELATED PCT APPLICATION NUMBER: PCT/US03/23075

By Authority of the  
COMMISSIONER OF PATENTS AND TRADEMARKS



H. L. JACKSON  
Certifying Officer

PRIORITY  
DOCUMENTSUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

07/24/02



07-25-02

605-482-5207-2407 A1PR2

Express Mail No.: EV 058 489 450 US

## COVER SHEET FOR PROVISIONAL APPLICATION FOR PATENT



Assistant Commissioner for Patents  
Box PROVISIONAL PATENT APPLICATION  
Washington, DC 20231

Sir:

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

		Docket Number	10589-023-888	Type a plus sign (+) inside this box -	+
INVENTOR(S) APPLICANT(S)					
LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)		
TITLE OF THE INVENTION (280 characters max)					
METHODS FOR IDENTIFYING SMALL MOLECULES THAT MODULATE PREMATURE TRANSLATION TERMINATION AND NONSENSE MEDIATED mRNA DECAY					
PENNIE & EDMONDS LLP					
CORRESPONDENCE ADDRESS		1000 19th Street, N.W., Suite 20583			
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification	Number of Pages	144	<input type="checkbox"/> Applicant claims small entity status, see 37 CFR §1.27		
<input checked="" type="checkbox"/> Drawing(s)	Number of Sheets	6	<input type="checkbox"/> Other (specify)		
METHOD OF PAYMENT (check one)					
<input type="checkbox"/> A check or money order is enclosed to cover the Provisional filing fees				ESTIMATED PROVISIONAL FILING FEE AMOUNT <input checked="" type="checkbox"/> \$160 <input type="checkbox"/> \$80	
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge the required filing fee to Deposit Account Number 16-1150					

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government  
 No.  Yes, the name of the U.S. Government agency and the Government contract number are. \_\_\_\_\_

Respectfully submitted,

Signature Laura A. Coruzzi  
 Laura A. Coruzzi  
 PENNIE & EDMONDS LLP

REGISTRATION NO  
(if appropriate)

30,742

Date

July 24, 2002

Additional inventors are being named on separately numbered sheets attached hereto

Total number of cover sheet pages

151

PROVISIONAL APPLICATION FILING ONLY

EXPRESS MAIL NO. EV 058 489 450 US  
ATTORNEY DOCKET NO. 10589-023-888

**METHODS FOR IDENTIFYING SMALL MOLECULES THAT  
MODULATE PREMATURE TRANSLATION TERMINATION  
AND NONSENSE MEDIATED mRNA DECAY**

---

5

**1. INTRODUCTION**

The present invention relates to a method for screening and identifying test compounds that modulate premature translation termination and/or nonsense-mediated messenger ribonucleic acid ("mRNA") decay by interacting with a preselected target ribonucleic acid ("RNA"). In particular, the present invention relates to identifying test compounds that bind to regions of the 28S ribosomal RNA ("tRNA") and analogs thereof. Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

20

**2. BACKGROUND OF THE INVENTION**

Protein synthesis encompasses the processes of translation initiation, elongation, and termination, each of which has evolved to occur with great accuracy and has the capacity to be a regulated step in the pathway of gene expression. Recent studies, including those suggesting that events at termination may regulate the ability of ribosomes to recycle to the start site of the same mRNA, have underscored the potential of termination to regulate other aspects of translation. The RNA triplets UAA, UAG, and UGA are non-coding and promote translational termination. Termination starts when one of the three termination codons enters the A site of the ribosome signaling the polypeptide chain release factors to bind and recognize the termination signal. Subsequently, the ester bond between the 3' nucleotide of the transfer RNA ("tRNA") located in the ribosome's P site and the nascent polypeptide chain is hydrolyzed, the completed polypeptide chain is released, and the ribosome subunits are recycled for another round of translation.

Nonsense-mediated mRNA decay is a surveillance mechanism that minimizes the translation and regulates the RNA stability of nonsense RNAs that contain chain termination mutations (see, e.g., Hentze & Kulozik, 1999, *Cell* 96:307-310; Culbertson, 1999, *Trends in Genetics* 15:74-80; Li & Wilkinson, 1998, *Immunity* 8:135-

141; and Ruiz-Echevarria *et al.*, 1996, Trends in Biological Sciences, 21:433-438). Chain termination mutations are caused when a base substitution or frameshift mutation changes a codon into a termination codon, *i.e.*, a premature stop codon that causes translational termination. In nonsense-mediated mRNA decay, mRNAs with premature stop codons are frequently subject to degradation. A truncated protein is produced as a result of the translation apparatus prematurely terminating at the stop codon.

5 Certain classes of known antibiotics have been characterized and found to interact with RNA. For example, the antibiotic thiostrepton binds tightly to a 60-mer from 10 ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, *Ann. Rev. Microbiol.* 43:207-233). Aminoglycosidic aminocyclitol 15 (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, *Nature* (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, *RNA* 1:95-101). In addition, 20 certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, *EMBO J.* 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain 25 aminoglycosides (Purohit *et al.*, 1994, *Nature* 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, *Nucleic Acids Res.* 21:4174-4179).

30 Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the 35 RRE RNA structural motif.

30 Aminoglycosides have also been found to promote nonsense suppression (see, *e.g.*, Bedwell *et al.*, 1997, *Nat. Med.* 3:1280-1284 and Howard *et al.*, 1996, *Nat. Med.* 2:467-469). Nonsense mutations cause approximately 10 to 30 percent of the individual cases of virtually all inherited diseases. Although nonsense mutations inhibit the synthesis of a full-length protein to one percent or less of wild-type levels, minimally boosting the expression levels of the full-length protein to between five and fifteen percent of normal

levels can eliminate or greatly reduce the severity of the disease. Clinical approaches that target the translation termination event to promote nonsense suppression have recently been described for model systems of cystic fibrosis and muscular dystrophy. Gentamicin is an aminoglycoside antibiotic that causes translational misreading and allowed the insertion of an amino acid at the site of the nonsense codon in models of cystic fibrosis, Hurlers Syndrome, and muscular dystrophy (see, e.g., Barton-Davis *et al.*, 1999, *J. Clin. Invest.* 104:375-381). These results strongly suggest that drugs that promote nonsense suppression by altering translation termination efficiency of a premature termination codon can be therapeutically valuable in the treatment of diseases caused by nonsense mutations.

10 Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

### **3. SUMMARY OF THE INVENTION**

15 The present invention relates to methods for identifying compounds that modulate translation termination and/or nonsense-mediated mRNA decay by identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. In particular, the present invention relates to identifying test compounds that bind to regions of the 28S rRNA and analogs thereof. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements (*i.e.*, regions of the 28S rRNA and analogs thereof) are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any 20 complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which 25 can be separated from the unbound beads and unbound target RNA in the liquid phase by a number of physical means, including, but not limited to, flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex 30 can then be identified by the label on the target RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound complexed 35

with the labeled RNA is then ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as 5 inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that 10 directly bind to 28S rRNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to 15 high-throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to 28S rRNA preferably under physiologic conditions. As a result, the compounds so 20 identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the 25 detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

25

### 3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not 30 limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, 28S rRNA and structural analogs thereof, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1

As used herein, a "library" refers to a plurality of test compounds with which 35 a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a

collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (e.g.,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^{3}\text{H}$ ), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (e.g., as commonly used in enzyme-linked immunoassays, e.g., alkaline phosphatase and horse radish peroxidase), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include an "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary ligands include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (e.g., oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer complexes, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may

emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucuronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound the test compound, can be easily identified.

#### 4. DESCRIPTION OF DRAWINGS

FIG. 1. The human 28S rRNA. Domains II and V are circled.

30 FIG. 2. Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain V of the 28S rRNA. 100 pmol of ribosomes were incubated with 100 $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

5 FIG. 3. Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain V of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

10 FIG. 4. Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain II (GTPase Center) of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

15 FIG. 5. Small molecules involved in nonsense suppression alter the chemical footprinting pattern of domain II of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with chemical modifying agents dimethyl sulfate (DMS) and kethoxal (KE). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

20 FIG. 6. A specific region of Domain II can compete for compound binding and prevents nonsense suppression *in vitro*. The *in vitro* nonsense suppression assay was performed using a luciferase construct with a UGA nonsense mutation. 0.1  $\mu$ M compound was present in the reaction to induce nonsense suppression. Competitor RNA corresponding to Domain II was added at the indicated concentrations (0, 1, 2.5, 5, 7.5, 10 pM) to titrate the small molecule and prevent nonsense suppression.

25

30

##### 5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. In particular, the present invention relates to identifying test compounds that bind to regions of

the 28S rRNA and analogs thereof. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the 5 present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which can be separated from the unbound target RNA in the liquid phase by a number of physical means, such as, but not limited to, flow cytometry, affinity chromatography, manual batch 10 mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex can then be identified by the label on the target RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound attached to the labeled RNA is then 15 ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay 20 for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are 25 ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding 30 assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to 28S rRNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for 35 modulating the function and/or activity of 28S rRNA.

Section 5.1 describes examples of 28S rRNA and analogs thereof that can be used as preselected target RNAs. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

### 5.1. 28S rRNA and Analogs Thereof

The ribosome is a 2.5-MDa ribonucleoprotein complex involved in the decoding of genetic material from mRNA to proteins. A combination of biophysical and biochemical analysis have provided three dimensional models of the ribosome as well as detailed analyses into the mechanism of the individual steps in translation (see, *e.g.*, Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716; Cate *et al.*, 1999, *Science* 285(5436):2095-2104; and Ban *et al.*, 2000, *Science*.289(5481): 905-920.).

The 28S rRNA is one of the ribosomal RNA components of the 60S subunit of eucaryotic ribosomes. The 28S rRNA sequences are conserved when expressed as mature rRNAs, although the 28S rRNA contains variable sequence tracts that are interspersed among conserved core sequences and lacking in the counterpart bacterial 23S rRNA (see, *e.g.*, Hancock & Dover, 1988, *Mol. Biol. Evol.* 5:377-391). A diagram of the 28S rRNA is presented in Figure 1, with domains II and V circled. As indicated in Figure 1, a GTPase center has been mapped to domain II and the peptidyl transferase center has been mapped to domain V.

Compounds that interact in these regions or modulate local changes within these domains of the ribosome (*e.g.*, alter base pairing interactions, base modification or modulate binding of trans-acting factors that bind to these regions) have the potential to modulate translation termination. These regions, *i.e.*, domains II and V are conserved from prokaryotes to eukaryotes, but the role of these regions in modulating translation termination has not been realized in eukaryotes. In bacteria, when a short RNA fragment, complementary to the *E. coli* 23S rRNA segment comprising nucleotides 735 to 766 (in

domain II), is expressed *in vivo*, suppression of UGA nonsense mutations, but not UAA or UAG, results (Chernyaeva *et al.*, 1999, *J Bacteriol* 181:5257-5262). Other regions of the 23S rRNA in *E. coli* have been implicated in nonsense suppression including the GTPase center in domain II (nt 1034-1120; Jemiolo *et al.*, 1995, *Proc. Nat. Acad. Sci.* 92:12309-12313).

5 Genetic mutations in bacteria have also identified rRNA mutations that either increase the level of frameshifting in the *trpE* or the suppression of a nonsense mutations in the *trpA* gene (reviewed in Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716). The frameshifting mutations mapped to domains IV and V of the 23S rRNA. 10 Disruption of the interaction of the CCA end of the tRNA with the peptidyl transferase center of the ribosome has been demonstrated to result in an increased translational error frequency (reviewed in Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716).

15 Regions of the 28S rRNA involved in frameshifting, nonsense mutation suppression, GTPase activity, or peptidyl transferase are attractive target RNAs to identify compounds that modulate premature translation termination and/or nonsense mediated decay. The interference of a test compound with one or more of these functions could potentially mediate translation termination by interfering with premature translation 20 termination. Without being bound by theory, a test compound could potentially mediate translation termination by causing readthrough of a premature translation codon, therefore allowing the synthesis of the full-length protein.

25 In a preferred embodiment, the target RNA comprises a region of 28S rRNA corresponding to domain II (see, *e.g.*, nucleotides 1310 to 2333 of accession number M11167) or domain V of 28S rRNA (see, *e.g.*, nucleotides 3859 to 4425 of accession number M11167) or an analog thereof. It will become apparent to one of skill in the art that an analog of the 28S rRNA has an analogous structure and function to native 28S rRNA. For example, an analog of human 28S rRNA includes, but is not limited to, a human 28S rRNA retropseudogene (see, *e.g.*, Wang *et al.*, 1997, *Gene* 196:105-111, Accession Number L20636). Regions corresponding to domain II or domain V of the 28S rRNA pseudogene 30 could be used as target RNAs in the present invention. In a preferred embodiment, the 28S rRNA is a human 28S rRNA, although the teachings of the present invention are applicable to mammals.

35 Synthesis of the target RNAs, *i.e.*, regions of 28S rRNA, can be performed by methods known to one of skill in the art (see, *e.g.*, Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York and Glover, D.M. (ed.), 1985, *DNA Cloning: A Practical Approach*,

MRL Press, Ltd., Oxford, U.K. Vol. I, II). In a preferred embodiment, the target RNAs are cloned as DNAs downstream of a promoter, such as but not limited to T7, T3, or Sp6 promoters, and *in vitro* transcribed with the corresponding polymerase. A detectable label can be incorporated into the *in vitro* transcribed RNA or alternatively, the target RNA is 5 end-labeled (see Section 5.2 *infra*). Alternatively, the target RNA can be amplified by polymerase chain reaction with a primer containing an RNA promoter and subsequently *in vitro* transcribed, as described in U.S. Patent No. 6,271,002, which is incorporated by reference in its entirety.

10

### 5.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled 15 with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

15

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (e.g., fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (e.g., 20 Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (e.g.,  $^3\text{H}$ ,  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{33}\text{P}$ , *etc.*), enzymes (e.g., horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal 25 gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. The label may be coupled directly or indirectly to a component of the detection assay (e.g., 30 the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (see, e.g., Caruthers *et al.*, U.S. Pat. Nos. 35 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, Genetic Engineering 4:1-17; *Users Manual*

Model 392 and 394 Polynucleotide Synthesizers, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, Biochemistry, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is 5 performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

10 The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated 15 intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate. Nucleoside addition is generally complete within 30 seconds. Next, a capping step is 20 performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the 25 oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate 30 protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

35 Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry, 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative

5 dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, e.g., Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

10 In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (e.g., a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled 15 ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated 20 strand synthesis.

20 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include, but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid 25 Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an 30 amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP- S and kinase, which reacts 35 specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

35 A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a

covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, e.g., a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag<sub>2</sub>S, ZnS, CdS, CdTe, Au, or TiO<sub>2</sub>. Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, Angew. Chem. Int. Ed. 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

### 5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds on solid supports. In all of the embodiments described

below, all of the libraries can be synthesized on solid supports or the compounds of the library can be attached to solid supports by linkers.

In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphoric acids and  $\alpha$ -amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; benzodiazepines; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, *J. Comb. Chem.* 3:477-517; Hall *et al.*, 2001, *ibid.* 3:125-150; Dolle, 2000, *ibid.* 2:383-433; Dolle, 1999, *ibid.* 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase

synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry.

5 Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 10 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

10 Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 15 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

15 In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or 20 solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid-phase synthesis of combinatorial compound libraries, liquid-phase synthesis does not require the use of 25 specialized protocols for monitoring the individual steps of a multistep solid-phase synthesis (Egner *et al.*, 1995, J.Org. Chem. 60:2652; Anderson *et al.*, 1995, J. Org. Chem. 60:2650; Fitch *et al.*, 1994, J. Org. Chem. 59:7955; Look *et al.*, 1994, J. Org. Chem. 49:7588; Metzger *et al.*, 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist *et al.*, 30 1994, Rapid Commun. Mass Spect. 8:77; Chu *et al.*, 1995, J. Am. Chem. Soc. 117:5419; Brummel *et al.*, 1994, Science 264:399; Stevanovic *et al.*, 1993, Bioorg. Med. Chem. Lett. 35 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, *Chem. Rev.* 97:41-448; Ohlmeyer *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, *Chem. Rev.* 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, doped polystyrene beads (as described by Fenniri *et al.*, 2000, *J. Am. Chem. Soc.* 123:8151-8152), alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden). In another embodiment, the solid support can be a magnetic bead coated with streptavidin, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway).

In one embodiment, the solid phase support is suitable for *in vivo* use, *i.e.*, it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestable.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after

synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, 5 photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. 10 Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

Dynamic combinatorial chemistry uses non-covalent interaction with a target 15 biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind 20 most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or 25 electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, 30 preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or 35 products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant

product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in its entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

10

#### 5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor is formed *in vitro*. For example, TK buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

25

The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that

30

35

is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

5        The solution further comprises a buffer, a combination of salts, and  
optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, Mops, 10 Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the types 15 of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

20        In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. Without being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl<sub>2</sub> stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in 25 the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

30        The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. 35 Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-

propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl  $\alpha$ -D-glucopyranoside, n-decyl  $\beta$ -D-maltopyranoside, n-dodecyl  $\beta$ -D-maltoside, n-octyl  $\beta$ -D-glucopyranoside, sorbitan esters, n-tetradecyl  $\beta$ -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxy polyethoxyethanol (NP-40), and tritons. 5 Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an 10 amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, *e.g.*, bovine serum 15 albumin ("BSA"), before contacting with to the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably 20 performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test 25 compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using any of the methods disclosed in Section 4.5 *infra*.

30

### **5.5. Separation Methods for Screening Test Compounds**

After the labeled target RNA is contacted with the library of test compounds immobilized on beads, the beads must then be separated from the unbound target RNA in the liquid phase. This can be accomplished by any number of physical means; *e.g.*, 35 sedimentation, centrifugation. Thereafter, a number of methods can be used to separate the library beads that are complexed with the labeled target RNA from uncomplexed beads in

order to isolate the test compound on the bead. Alternatively, mass spectroscopy and NMR spectroscopy can be used to simultaneously identify and separate beads complexed to the labeled target RNA from uncomplexed beads.

5

### 5.5.1. Flow Cytometry

In a preferred embodiment, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. Flow cytometers for sorting and examining biological cells are well known in the art; this technology can be applied to separate the labeled library beads from unlabeled beads. Known flow cytometers are described, for example, in U.S. Patent Nos. 4,347,935; 5,464,581; 5,483,469; 5,602,039; 5,643,796; and 6,211,477; the entire contents of which are incorporated by reference herein. Other known flow cytometers are the FACS Vantage™ system manufactured by Becton Dickinson and Company, and the COPAS™ system manufactured by Union Biometrica.

15 A flow cytometer typically includes a sample reservoir for receiving a biological sample. The biological sample contains particles (hereinafter referred to as "beads") that are to be analyzed and sorted by the flow cytometer. Beads are transported from the sample reservoir at high speed (>100 beads/second) to a flow cell in a stream of liquid "sheath" fluid. High-frequency vibrations of a nozzle that directs the stream to the

20 flow cell causes the stream to partition and form ordered droplets, with each droplet containing a single bead. Physical properties of beads can be measured as they intersect a laser beam within the cytometer flow cell. As beads move one by one through the interrogation point, they cause the laser light to scatter and fluorescent molecules on the labeled beads (*i.e.*, beads complexed with labeled target RNA) become excited.

25 Alternatively, if the target nucleic acid is labeled with an inorganic nanoparticle, the beads complexed with bound target nucleic acid can be distinguished not only by unique fluorescent properties but also on the basis of spectrometric properties (*e.g.* including but not limited to increased optical density due to the reduction of Ag<sup>+</sup> ions in the presence of gold nanoparticles (see, *e.g.*, Taton *et al.* *Science* 2000, 289: 1757-1760)).

30 An appropriate detection system consisting of photomultiplier tubes, photodiodes or other devices for measuring light are focused onto the interrogation point where the properties are measured. In so doing, information regarding particle size (light scatter) and complex formation (fluorescence intensity) is obtained. Particles with the desired physical properties are then sorted by a variety of physical means. In one embodiment, the beads are sorted by an electrostatic method. To sort beads by an electrostatic method, the droplets containing the beads with the desired physical properties

)

are electrically charged and deflected from the trajectory of uncharged droplets as they pass through an electrostatic field formed by two deflection plates held constant at a high electrical potential difference. In another embodiment, the beads are sorted by an air-diverting method. To sort beads by an air-diverting method, the droplets containing the beads with the desired physical properties are deflected from their trajectory by a focused stream of forced air. Both of these embodiments cause the trajectory of beads with the desired physical properties to become changed, thereby sorting them from other beads. Accordingly, the beads complexed to the labeled target RNA can be collected in an appropriate collecting vessel.

Thus, in one embodiment of the present invention, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. In a preferred embodiment, the target nucleic acid is labeled with a fluorescent label and the complexed and non-complexed target nucleic acids are separated by fluorescence activated cell sorting ("FACS"). Such methods are well known to one of skill in the art.

#### 5.5.2. Affinity Chromatography

In another embodiment of the invention, the target RNA can be labeled with biotin, an antigen, or a ligand. Library beads complexed to the target RNA can be separated from uncomplexed beads using affinity techniques designed to capture the labeled moiety on the target RNA. For example, a solid support, such as but not limited to, a column or a well in a microwell plate coated with avidin/streptavidin, an antibody to the antigen, or a receptor for the ligand can be used to capture or immobilize the labeled beads. Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents. See, *e.g.*, International Patent Publication WO/0146461, the contents of which are hereby incorporated by reference. The unbound library beads can be removed after the binding reaction by washing the solid phase. If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by preferably, but not limited to, enzymatic or chemical (*e.g.*, alkaline hydrolysis) degradation. The library beads bound to the solid phase can then be eluted with any solution that disrupts the binding between the labeled target RNA and the solid phase. Such

solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

5 In one embodiment, the library of test compounds can be prepared on magnetic beads, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway). The magnetic bead library can then be mixed with the labeled target RNA under conditions that allow binding to occur. The separation of the beads from unbound target RNA in the liquid phase can be accomplished using a magnet. After removal of the magnetic field, the bead complexed to the labeled RNA may be separated from uncomplexed library beads via the 10 label used on the target RNA; e.g., biotinylated target RNA can be captured by avidin/streptavidin; target RNA labeled with antigen can be captured by the appropriate antibody; target RNA labeled with ligand can be captured using the appropriate immobilized receptor. The captured library bead can then be eluted with any solution that 15 disrupts the binding between the labeled target RNA and the immobilized surface. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not 20 limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents. See, e.g., International Patent Publication WO/0146461, the contents of which are hereby 25 incorporated by reference. If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U<sub>2</sub>, CL<sub>3</sub>, T<sub>1</sub>, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (e.g. nickel(II), cobalt(II), or iron(II)) oxidative 30 cleavage.

35 In another embodiment, the preselected target RNA can be labeled with a heavy metal tag and incubated with the library beads to allow binding of the test compounds to the target RNA. The separation of the labeled beads from unlabeled beads can be accomplished using a magnetic field. After removal of the magnetic field, the test compound can be eluted with any solution that disrupts the binding between the preselected

target RNA and the test compound. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

5

### 5.5.3. Manual Batch

In one embodiment, a manual "batch" mode is used for separating complexed beads. To explore a bead-based library within a reasonable time period, the primary screens should be operated with sufficient throughput. To do this, the target 10 nucleic acid is labeled with a dye and then incubated with the combinatorial library. An advantage of such an assay is the fast identification of active library beads by color change. In the lower concentrations of the dye-labeled target molecule, only those library beads that bind the target molecules most tightly are detected because of higher local concentration of the dye. When washed and plated into a liquid monolayer, colored beads are easily 15 separated from non-colored beads with the aid of a dissecting microscope. One of the problems associated with this method could be the interaction between the red dye and library substrates. Control experiments using the dye alone and dye attached to mutant RNA sequences with the libraries are performed to eliminate this possibility.

20

### 5.5.4. Suspension of Beads in Electric Fields

25

In another embodiment of the invention, library beads bound to the target RNA can be separated from unbound beads on the basis of the altered charge properties due to RNA binding. In a preferred embodiment of this technique, beads are separated from unbound nucleic acid and suspended, preferably but not only, in the presence of an electric field where the bound RNA causes the beads bound to the target RNA to migrate toward the anode, or positive, end of the field.

30

Beads can be preferentially suspended in solution as a colloidal suspension with the aid of detergents or surfactants. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid, carboxymethylcellulose, carrageenan, and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alky amidoalkyl betaines, 35 N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl  $\alpha$ -D-glucopyranoside, n-decyl-D-maltopyranoside,

n-dodecyl -D-maltoside, n-octyl -D-glucopyranoside, sorbitan esters, n-tetradecyl -D-maltoside and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol.

Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents.

If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U<sub>2</sub>, CL<sub>3</sub>, T<sub>1</sub>, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (e.g. nickel(II), cobalt(II), or iron(II)) oxidative cleavage.

### 5.5.5. Microwave Spectroscopy

In another embodiment, the complexed beads are separated from uncomplexed beads by microwave spectroscopy. For example, as described in U.S. Patent Nos. 6,395,480; 6,376,258; 6,368,795; 6,340,568; 6,338,968; 6,287,874; and 6,287,776 to Hefti, the disclosures of which are hereby incorporated by reference, the unique dielectric properties of molecules and binding complexes, such as hybridization complexes formed between a nucleic acid probe and a nucleic acid target, molecular binding events, and protein/ligand complexes, result in varying microwave spectra which can be measured. The molecule's dielectric properties can be observed by coupling a test signal to the molecule and observing the resulting signal. When the test signal excites the molecule at a frequency within the molecule's dispersion regime, especially at a resonant frequency, the molecule will interact strongly with the signal, and the resulting signal will exhibit dramatic variations in its measured amplitude and phase, thereby generating a unique signal response. This response can be used to detect and identify the bound molecular structure. In addition, because most molecules will exhibit different dispersion properties over the same or

different frequency bands, each generates a unique signal response which can be used to identify the molecular structure.

### 5 5.6. Methods for Identifying Test Compounds

If the library is a peptide or nucleic acid library, the sequence of the test compound on the isolated bead can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

#### 10 5.6.1. Mass Spectrometry

Mass spectrometry (e.g., electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound on the isolated bead.

15 MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, *J. Biol. Chem.* 272:32084-32091). However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the 20 MALDI process.

20 ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, *Anal. Chem.* 69:5130-5135).

25 Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, *Anal. Chem.* 71:3436-3440; Griffey *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:10129-10133). As true for all of the mass spectrometry 30 methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

35 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery

of a consensus structure of a test compound that specifically binds to a preselected target RNA.

In a preferred embodiment, the structure of the test compound is determined by time of flight mass spectroscopy ("TOF-MS"). In time of flight methods of mass spectrometry, charged (ionized) molecules are produced in a vacuum and accelerated by an electric field into a time of flight tube or drift tube. The velocity to which the molecules may be accelerated is proportional to the accelerating potential, proportional to the charge of the molecule, and inversely proportional to the square of the mass of the molecule. The charged molecules travel, *i.e.*, "drift" down the TOF tube to a detector. The time taken for the molecules to travel down the tube may be interpreted as a measure of their molecular weight. Time-of-flight mass spectrometers have been developed for all of the major ionization techniques such as, but limited to, electron impact ("EI"), infrared laser desorption ("IRLD"), plasma desorption ("PD"), fast atom bombardment ("FAB"), secondary ion mass spectrometry ("SIMS"), matrix-assisted laser desorption/ionization ("MALDI"), and electrospray ionization ("ESI").

#### **5.6.2. NMR Spectroscopy**

NMR spectroscopy can be used for elucidating the structure of the test compound on the isolated bead. NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

#### **5.6.3. Edman Degradation**

In an embodiment wherein the library is a peptide library or a derivative thereof, Edman degradation can be used to determine the structure of the test compound. In one embodiment, a modified Edman degradation process is used to obtain compositional

tags for proteins, which is described in U.S. Patent No. 6,277,644 to Farnsworth *et al.*, which is hereby incorporated by reference in its entirety. The Edman degradation chemistry is separated from amino acid analysis, circumventing the serial requirement of the conventional Edman process. Multiple cycles of coupling and cleavage are performed prior to extraction and compositional analysis of amino acids. The amino acid composition information is then used to search a database of known protein or DNA sequences to identify the sample protein. An apparatus for performing this method comprises a sample holder for holding the sample, a coupling agent supplier for supplying at least one coupling agent, a cleavage agent supplier for supplying a cleavage agent, a controller for directing the sequential supply of the coupling agents, cleavage agents, and other reagents necessary for performing the modified Edman degradation reactions, and an analyzer for analyzing amino acids.

In another embodiment, the method can be automated as described in U.S. Patent No. 5,565,171 to Dovichi *et al.*, which is hereby incorporated by reference in its entirety. The apparatus includes a continuous capillary connected between two valves that control fluid flow in the capillary. One part of the capillary forms a reaction chamber where the sample may be immobilized for subsequent reaction with reagents supplied through the valves. Another part of the capillary passes through or terminates in the detector portion of an analyzer such as an electrophoresis apparatus, liquid chromatographic apparatus or mass spectrometer. The apparatus may form a peptide or protein sequencer for carrying out the Edman degradation reaction and analyzing the reaction product produced by the reaction. The protein or peptide sequencer includes a reaction chamber for carrying out coupling and cleavage on a peptide or protein to produce derivatized amino acid residue, a conversion chamber for carrying out conversion and producing a converted amino acid residue and an analyzer for identifying the converted amino acid residue. The reaction chamber may be contained within one arm of a capillary and the conversion chamber is located in another arm of the capillary. An electrophoresis length of capillary is directly capillary coupled to the conversion chamber to allow electrophoresis separation of the converted amino acid residue as it leaves the conversion chamber. Identification of the converted amino acid residue takes place at one end of the electrophoresis length of the capillary.

#### 5.6.4. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, must are observable only with one or the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

In one embodiment of the method, test compounds are synthesized on 10 polystyrene beads doped with chemically modified styrene monomers such that each resulting bead has a characteristic pattern of absorption lines in the vibrational (IR or Raman) spectrum, by methods including but not limited to those described by Fenniri *et al.*, 2000, J. Am. Chem. Soc. 123:8151-8152. Using methods of split-pool synthesis familiar to one of skill in the art, the library of compounds is prepared so that the spectroscopic pattern 15 of the bead identifies one of the components of the test compound on the bead. Beads that have been separated according to their ability to bind target RNA can be identified by their vibrational spectrum. In one embodiment of the method, appropriate sorting and binning of the beads during synthesis then allows identification of one or more further components of the test compound on any one bead. In another embodiment of the method, partial 20 identification of the compound on a bead is possible through use of the spectroscopic pattern of the bead with or without the aid of further sorting during synthesis, followed by partial resynthesis of the possible compounds aided by doped beads and appropriate sorting during synthesis.

In another embodiment, the IR or Raman spectra of test compounds are 25 examined while the compound is still on a bead, preferably, or after cleavage from bead, using methods including but not limited to photochemical, acid, or heat treatment. The test compound can be identified by comparison of the IR or Raman spectral pattern to spectra previously acquired for each test compound in the combinatorial library.

30

#### 5.6.5. Microwave Spectroscopy

In another embodiment, the microwave spectra of a test compound can be used to elucidate the structure of the test compound. For example, as described in U.S. Patent Nos. 6,395,480; 6,376,258; 6,368,795; 6,340,568; 6,338,968; 6,287,874; and 35 6,287,776 to Hefti, the disclosures of which are hereby incorporated by reference, the unique dielectric properties of molecules and binding complexes, such as hybridization complexes formed between a nucleic acid probe and a nucleic acid target, molecular

binding events, and protein/ligand complexes, result in varying microwave spectra which can be measured. The molecule's dielectric properties can be observed by coupling a test signal to the molecule and observing the resulting signal. When the test signal excites the molecule at a frequency within the molecule's dispersion regime, especially at a resonant frequency, the molecule will interact strongly with the signal, and the resulting signal will exhibit dramatic variations in its measured amplitude and phase, thereby generating a unique signal response. This response can be used to detect and identify the bound molecular structure. In addition, because most molecules will exhibit different dispersion properties over the same or different frequency bands, each generates a unique signal response which can be used to identify the molecular structure.

### 5.7. Secondary Biological Screens

10 The test compounds identified in the nonsense suppression assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using an *in vitro* transcribed RNA from the gene with the premature translation termination codon and *in vitro* translating that RNA in a cell-free translation extract. The activity of the lead compound in the *in vitro* translation mixture can be determined by any method that measures increased expression of the full-length gene, *i.e.*, past the premature termination codon. For example, expression of a functional protein from the full-length gene (*e.g.*, a reporter gene) can be measured to determine the effect of the lead compound on premature translation termination and/or nonsense-mediated mRNA decay in an *in vitro* system.

15 In addition, the lead compound can be tested in a host cell engineered to contain the RNA with the premature translation termination codon controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the RNA with the premature translation termination codon. Compounds that modulate premature translation termination and/or nonsense-mediated mRNA decay *in vivo* will result in increased expression of the full-length gene, *i.e.*, past the premature termination codon. Alternatively, a phenotypic or physiological readout can be 20 used to assess activity of the target RNA with the premature translation termination codon in the presence and absence of the lead compound. Both the *in vitro* and *in vivo* nonsense suppression assays used herein and as described in International Patent Publication WO 01/44516, which is incorporated by reference in its entirety, can be used to identify lead compounds can also be used to determine an EC<sub>50</sub> for the lead compounds.

25 The test compounds identified in the nonsense suppression assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity

using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the RNA with the premature translation termination codon controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the RNA with the premature translation termination codon.

5 Compounds that modulate premature translation termination and/or nonsense-mediated mRNA decay *in vivo* will result in increased expression of the full-length gene, *i.e.*, past the premature termination codon. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA with the premature translation termination codon 10 in the presence and absence of the lead compound. Both the *in vitro* and *in vivo* nonsense suppression assays used herein and as described in International Patent Publication WO 01/44516, which is incorporated by reference in its entirety, can be used to identify lead compounds can also be used to determine an EC<sub>50</sub> for the lead compounds.

15 Animal model systems can also be used to demonstrate the safety and efficacy of the lead compounds identified in the nonsense suppression assays described above. The lead compounds identified in the nonsense suppression assay can then be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional 20 readout system, such as a transgenic mouse.

25 Examples of animal models for cystic fibrosis include, but are not limited to, cftr(-/-) mice (see, *e.g.*, Freedman *et al.*, 2001, *Gastroenterology* 121(4):950-7), cftr(tm1HGU/tm1HGU) mice (see, *e.g.*, Bernhard *et al.*, 2001, *Exp Lung Res* 27(4):349-66), CFTR-deficient mice with defective cAMP-mediated Cl(-) conductance (see, *e.g.*, Stotland *et al.*, 2000, *Pediatr Pulmonol* 30(5):413-24), and C57BL/6-  
Cftr(m1UNC)/Cftr(m1UNC) knockout mice (see, *e.g.*, Stotland *et al.*, 2000, *Pediatr Pulmonol* 30(5):413-24).

30 Examples of animal models for muscular dystrophy include, but are not limited to, mouse, hamster, cat, dog, and *C. elegans*. Examples of mouse models for muscular dystrophy include, but are not limited to, the dy-/- mouse (see, *e.g.*, Connolly *et al.*, 2002, *J Neuroimmunol* 127(1-2):80-7), a muscular dystrophy with myositis (mdm) mouse mutation (see, *e.g.*, Garvey *et al.*, 2002, *Genomics* 79(2):146-9), the mdx mouse (see, *e.g.*, Nakamura *et al.*, 2001, *Neuromuscul Disord* 11(3):251-9), the utrophin-dystrophin knockout (dko) mouse (see, *e.g.*, Nakamura *et al.*, 2001, *Neuromuscul Disord* 11(3):251-9), the dy/dy mouse (see, *e.g.*, Dubowitz *et al.*, 2000, *Neuromuscul Disord* 10(4-5):292-8), the mdx(Cv3) mouse model (see, *e.g.*, Pillers *et al.*, 1999, *Laryngoscope*

109(8):1310-2), and the myotonic ADR-MDX mutant mice (see, e.g., Kramer *et al.*, 1998, Neuromuscul Disord 8(8):542-50). Examples of hamster models for muscular dystrophy include, but are not limited to, sarcoglycan-deficient hamsters (see, e.g., Nakamura *et al.*, 5 2001, Am J Physiol Cell Physiol 281(2):C690-9) and the BIO 14.6 dystrophic hamster (see, e.g., Schlenker & Burbach, 1991, J Appl Physiol 71(5):1655-62). An example of a feline model for muscular dystrophy includes, but is not limited to, the hypertrophic feline muscular dystrophy model (see, e.g., Gaschen & Burgunder, 2001, Acta Neuropathol (Berl) 101(6):591-600). Canine models for muscular dystrophy include, but are not limited to, golden retriever muscular dystrophy (see, e.g., Fletcher *et al.*, 2001, Neuromuscul Disord 10 11(3):239-43) and canine X-linked muscular dystrophy (see, e.g., Valentine *et al.*, 1992, Am J Med Genet 42(3):352-6). Examples of *C. elegans* models for muscular dystrophy are described in Chamberlain & Benian, 2000, Curr Biol 10(21):R795-7 and Culette & Sattelle, 2000, Hum Mol Genet 9(6):869-77.

15 Examples of animal models for familial hypercholesterolemia include, but are not limited to, mice lacking functional LDL receptor genes (see, e.g., Aji *et al.*, 1997, Circulation 95(2):430-7), Yoshida rats (see, e.g., Fantappie *et al.*, 1992, Life Sci 50(24):1913-24), the JCR:LA-cp rat (see, e.g., Richardson *et al.*, 1998, Atherosclerosis 138(1):135-46), swine (see, e.g., Hasler-Rapacz *et al.*, 1998, Am J Med Genet 76(5):379- 20 86), and the Watanabe heritable hyperlipidaemic rabbit (see, e.g., Tsutsumi *et al.*, 2000, Arzneimittelforschung 50(2):118-21; Harsch *et al.*, 1998, Br J Pharmacol 124(2):227-82; and Tanaka *et al.*, 1995, Atherosclerosis 114(1):73-82).

25 An example of an animal model for human cancer in general includes, but is not limited to, spontaneously occurring tumors of companion animals (see, e.g., Vail & MacEwen, 2000, Cancer Invest 18(8):781-92). Examples of animal models for lung cancer include, but are not limited to, lung cancer animal models described by Zhang & Roth (1994, In Vivo 8(5):755-69) and a transgenic mouse model with disrupted p53 function (see, e.g., Morris *et al.*, 1998, J La State Med Soc 150(4):179-85). An example of an animal model for breast cancer includes, but is not limited to, a transgenic mouse that overexpresses cyclin D1 (see, e.g., Hosokawa *et al.*, 2001, Transgenic Res 10(5):471-8). 30 An example of an animal model for colon cancer includes, but is not limited to, a TCRbeta and p53 double knockout mouse (see, e.g., Kado *et al.*, 2001, Cancer Res 61(6):2395-8). Examples of animal models for pancreatic cancer include, but are not limited to, a metastatic model of Panc02 murine pancreatic adenocarcinoma (see, e.g., Wang *et al.*, 2001, Int J Pancreatol 29(1):37-46) and nu-nu mice generated in subcutaneous pancreatic tumours (see, e.g., Ghaneh *et al.*, 2001, Gene Ther 8(3):199-208). Examples of animal models for 35 )

non-Hodgkin's lymphoma include, but are not limited to, a severe combined immunodeficiency ("SCID") mouse (see, e.g., Bryant *et al.*, 2000, *Lab Invest* 80(4):553-73) and an IgHmu-HOX11 transgenic mouse (see, e.g., Hough *et al.*, 1998, *Proc Natl Acad Sci USA* 95(23):13853-8). An example of an animal model for esophageal cancer includes, but is not limited to, a mouse transgenic for the human papillomavirus type 16 E7 oncogene (see, e.g., Herber *et al.*, 1996, *J Virol* 70(3):1873-81). Examples of animal models for colorectal carcinomas include, but are not limited to, Apc mouse models (see, e.g., Fodde & Smits, 2001, *Trends Mol Med* 7(8):369-73 and Kuraguchi *et al.*, 2000, *Oncogene* 19(50):5755-63). An example of an animal model for neurofibromatosis includes, but is not limited to, mutant NF1 mice (see, e.g., Cichowski *et al.*, 1996, *Semin Cancer Biol* 7(5):291-8). Examples of animal models for retinoblastoma include, but are not limited to, transgenic mice that express the simian virus 40 T antigen in the retina (see, e.g., Howes *et al.*, 1994, *Invest Ophthalmol Vis Sci* 35(2):342-51 and Windle *et al.*, 1990, *Nature* 343(6259):665-9) and inbred rats (see, e.g., Nishida *et al.*, 1981, *Curr Eye Res* 1(1):53-5 and Kobayashi *et al.*, 1982, *Acta Neuropathol (Berl)* 57(2-3):203-8). Examples of animal models for Wilm's tumor include, but are not limited to, a WT1 knockout mouse (see, e.g., Scharnhorst *et al.*, 1997, *Cell Growth Differ* 8(2):133-43), a rat subline with a high incidence of nephroblastoma (see, e.g., Mesfin & Breech, 1996, *Lab Anim Sci* 46(3):321-6), and a Wistar/Furth rat with Wilms' tumor (see, e.g., Murphy *et al.*, 1987, *Anticancer Res* 7(4B):717-9).

Examples of animal models for retinitis pigmentosa include, but are not limited to, the Royal College of Surgeons ("RCS") rat (see, e.g., Vollrath *et al.*, 2001, *Proc Natl Acad Sci USA* 98(22):12584-9 and Hanitzsch *et al.*, 1998, *Acta Anat (Basel)* 162(2-3):119-26), a rhodopsin knockout mouse (see, e.g., Jaissle *et al.*, 2001, *Invest Ophthalmol Vis Sci* 42(2):506-13), and Wag/Rij rats (see, e.g., Lai *et al.*, 1980, *Am J Pathol* 98(1):281-4).

Examples of animal models for cirrhosis include, but are not limited to, CCl<sub>4</sub>-exposed rats (see, e.g., Kloehn *et al.*, 2001, *Horm Metab Res* 33(7):394-401) and rodent models instigated by bacterial cell components or colitis (see, e.g., Vierling, 2001, *Best Pract Res Clin Gastroenterol* 15(4):591-610).

Examples of animal models for hemophilia include, but are not limited to, rodent models for hemophilia A (see, e.g., Reipert *et al.*, 2000, *Thromb Haemost* 84(5):826-32; Jarvis *et al.*, 1996, *Thromb Haemost* 75(2):318-25; and Bi *et al.*, 1995, *Nat Genet* 10(1):119-21), canine models for hemophilia A (see, e.g., Gallo-Penn *et al.*, 1999, *Hum Gene Ther* 10(11):1791-802 and Connelly *et al.*, 1998, *Blood* 91(9):3273-81), murine

models for hemophilia B (see, e.g., Snyder *et al.*, 1999, *Nat Med* 5(1):64-70; Wang *et al.*, 1997, *Proc Natl Acad Sci USA* 94(21):11563-6; and Fang *et al.*, 1996, *Gene Ther* 3(3):217-22), canine models for hemophilia B (see, e.g., Mount *et al.*, 2002, *Blood* 99(8):2670-6; Snyder *et al.*, 1999, *Nat Med* 5(1):64-70; Fang *et al.*, 1996, *Gene Ther* 3(3):217-22); and Kay *et al.*, 1994, *Proc Natl Acad Sci USA* 91(6):2353-7), and a rhesus macaque model for hemophilia B (see, e.g., Lozier *et al.*, 1999, *Blood* 93(6):1875-81).

5 Examples of animal models for von Willebrand disease include, but are not limited to, an inbred mouse strain RIIS/J (see, e.g., Nichols *et al.*, 1994, 83(11):3225-31 and Sweeney *et al.*, 1990, 76(11):2258-65), rats injected with botrocetin (see, e.g., Sanders *et al.*, 1988, *Lab Invest* 59(4):443-52), and porcine models for von Willebrand disease (see, e.g., Nichols *et al.*, 1995, *Proc Natl Acad Sci USA* 92(7):2455-9; Johnson & Bowie, 1992, *J Lab Clin Med* 120(4):553-8); and Brinkhous *et al.*, 1991, *Mayo Clin Proc* 66(7):733-42).

10 Examples of animal models for  $\beta$ -thalassemia include, but are not limited to, murine models with mutations in globin genes (see, e.g., Lewis *et al.*, 1998, *Blood* 91(6):2152-6; Raja *et al.*, 1994, *Br J Haematol* 86(1):156-62; Popp *et al.*, 1985, 445:432-44; and Skow *et al.*, 1983, *Cell* 34(3):1043-52).

15 Examples of animal models for kidney stones include, but are not limited to, genetic hypercalciuric rats (see, e.g., Bushinsky *et al.*, 1999, *Kidney Int* 55(1):234-43 and Bushinsky *et al.*, 1995, *Kidney Int* 48(6):1705-13), chemically treated rats (see, e.g., Grases *et al.*, 1998, *Scand J Urol Nephrol* 32(4):261-5; Burgess *et al.*, 1995, *Urol Res* 23(4):239-42; Kumar *et al.*, 1991, *J Urol* 146(5):1384-9; Okada *et al.*, 1985, *Hinyokika Kiyo* 31(4):565-77; and Bluestone *et al.*, 1975, *Lab Invest* 33(3):273-9), hyperoxaluric rats (see, e.g., Jones *et al.*, 1991, *J Urol* 145(4):868-74), pigs with unilateral retrograde flexible 20 nephroscopy (see, e.g., Seifmah *et al.*, 2001, 57(4):832-6), and rabbits with an obstructed nephroscopy (see, e.g., Itatani *et al.*, 1979, *Invest Urol* 17(3):234-40).

25 Examples of animal models for ataxia-telangiectasia include, but are not limited to, murine models of ataxia-telangiectasia (see, e.g., Barlow *et al.*, 1999, *Proc Natl Acad Sci USA* 96(17):9915-9 and Inoue *et al.*, 1986, *Cancer Res* 46(8):3979-82).

30 Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling 35 (cf. Lamothe, *et al.* 1997, *J. Med. Chem.* 40: 3542; Mottola *et al.* 1996, *J. Med. Chem.* 39:

285; Beusen *et al.* 1995, *Biopolymers* 36: 181; P. Fossa *et al.* 1998, *Comput. Aided Mol. Des.* 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, *J. Med. Chem.* 42: 4122; Barreca *et al.* 1999 *Bioorg. Med. Chem.* 7: 2283; Kroemer *et al.* 1995, *J. Med. Chem.* 38: 4917; Schaal *et al.* 2001, *J. Med. Chem.* 44: 155; Buolamwini & Assefa 2002, *J. Mol. Chem.* 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, *J. Med. Chem.* 44: 1134; Faaland *et al.* 2000, *Biochem. Cell. Biol.* 78: 415; Silvestri *et al.* 2000, *Bioorg. Med. Chem.* 8: 2305; J. Lee *et al.* 2001, *Bioorg. Med. Chem.* 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described by Zuker *et al.* *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology* pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 *J. Mol. Biol.* 288: 911-940); RNAMotif (Macke *et al.* 2001, *Nucleic Acids Res.* 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, *Monatsh. Chem.* 125: 167-188).

15 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, *Acta Pharmaceutical Fennica* 97:159-166; Ripka, 1998, *New Scientist* 54-57; McKinaly & Rossmann, 1989, *Annu. Rev. Pharmacol. Toxicol.* 29:111-122; Perry & Davies, *QSAR: Quantitative Structure-Activity Relationships in Drug Design* pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, *Proc. R. Soc. Lond.* 236:125-140 and 141-162; Askew *et al.*, 1989, *J. Am. Chem. Soc.* 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (*e.g.*, Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (*e.g.*, Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (*e.g.*, QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (*e.g.* RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, 20 proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (*e.g.*, Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, *Adv. Drug Delivery Rev.* 23: 3-25). Most importantly, 25 they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the 30 35

synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

5

### **5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease**

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is associated with premature translation termination and/or nonsense-mediated mRNA decay. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with premature translation termination and/or nonsense-mediated mRNA decay.

15 In a preferred embodiment, it is first determined that the patient is suffering from a disease associated with premature translation termination and/or nonsense-mediated mRNA decay. In a preferred embodiment, the DNA of the patient can be sequenced or subject to Southern blot, polymerase chain reaction ("PCR"), use of the Short Tandem Repeat ("STR"), or polymorphic length restriction fragments ("RFLP") analysis to determine if a nonsense mutation is present in the DNA of the patient. Alternatively, it can be determined if altered levels of the protein with the nonsense mutation are expressed in the patient by western blot or other immunoassays. Such methods are well known to one of skill in the art.

20 25 In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, e.g., stabilization of a discernible symptom, physiologically, e.g., stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

30 35 In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with a disease caused and/or associated with nonsense suppression. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a

patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

5        In a preferred embodiment, the compounds identified using the methods of the present invention are used to treat or prevent a disease caused by one or more nonsense mutations. Examples of diseases caused by nonsense mutations include, but are not limited to, cystic fibrosis, muscular dystrophy, heart disease, lung cancer, breast cancer, colon cancer, pancreatic cancer, non-Hodgkin's lymphoma, ovarian cancer, esophageal cancer, 10      colorectal carcinomas, neurofibromatosis, retinoblastoma, Wilm's tumor, retinitis pigmentosa, collagen disorders, cirrhosis, Tay-Sachs disease, blood disorders, kidney stones, and ataxia-telangiectasia. Genes that contain one or more nonsense mutations that are potentially involved in causing disease are presented in table form according to 15      chromosome location in Example 9 *infra*.

15        When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus 20      injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable 25      salts thereof.

25        Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or 30      topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

30        In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by 35

means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

5 In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

10 Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

15 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

20 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, e.g., Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507; Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see 25 *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science* 228:190; During *et al.*, 1989, *Ann. Neurol.* 25:351; Howard *et al.*, 1989, *J. Neurosurg.* 30 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

35 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a

pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

In a specific embodiment, the term "pharmaceutically acceptable" means 5 approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, 10 vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the 15 pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include 20 excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering 25 agents.

Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other 25 form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see e.g., U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable 30 salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered 35 compositions may contain one or more agents, for example, sweetening agents such as

fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained 5 action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by 10 the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can 15 include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise 20 sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous 25 administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle 30 containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that 35 will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are

generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 5 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 10 95% active ingredient by weight.

15 Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

20 Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived 25 from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

30 The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

35 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphonic acids and  $\alpha$ -amino phosphinic acids,

5 or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

**6. EXAMPLE: COMPOUNDS THAT MODULATE TRANSLATION TERMINATION BIND SPECIFIC REGIONS OF 28S rRNA**

10 Data is presented in this Example that demonstrates that specific regions of the 28S rRNA are involved in modulating translation termination in mammalian cells. Compounds that interact in these regions or modulate local changes within these regions of the ribosome (e.g., alter base pairing interactions, base modification or modulate binding of trans-acting factors that bind to these regions) have the potential to modulate translation 15 termination. These regions are conserved from prokaryotes to eukaryotes, but the role of these regions in modulating translation termination has not been realized in eukaryotes. In bacteria, when a short RNA fragment, complementary to the *E. coli* 23S rRNA segment comprising nucleotides 735 to 766 (in domain II), is expressed *in vivo*, suppression of UGA nonsense mutations, but not UAA or UAG, results (Chernyaeva *et al.*, 1999, *J Bacteriol* 20 181:5257-5262). Other regions of the 23S rRNA in *E. coli* have been implicated in 20 nonsense suppression including the GTPase center in domain II (nt 1034-1120; Jemiolo *et al.*, 1995, *Proc. Nat. Acad. Sci.* 92:12309-12313).

**6.1. Materials and Methods**

25 **6.1.1. Small Molecules Involved in Modulating Translation Termination**  
 Small molecules involved in modulating translation termination, *i.e.*, nonsense suppression, were used in the footprinting experiments presented in Figures 2 to 6 and are listed as Compound A (molecular formula  $C_{19}H_{21}NO_4$ ), Compound B (molecular 30 formula  $C_{19}H_{21}N_2O_5$ ), Compound C (molecular formula  $C_{12}H_{15}N_5O$ ), Compound D (molecular formula  $C_{23}H_{15}O_3Br$ ), Compound E (molecular formula  $C_{19}H_{21}NO_4$ ), Compound F, Compound G (molecular formula  $C_{12}H_{15}N_5O$ ), Compound H (molecular formula  $C_{23}H_{15}NO_5$ ), Compound I (molecular formula  $C_{23}H_{15}NO_5$ ), Compound J, and Compound K.

### 6.1.2. Preparation of a Translation Extract from HeLa cells

HeLa S3 cells were grown to a density of  $10^6$  cells/ml in DMEM; 5%CO<sub>2</sub>, 10%FBS, 1X P/S in a spinner flask. Cells were harvested by spinning at 1000Xg. Cells were washed twice with phosphate buffered saline. The cell pellet was on ice for 12-24 hours before proceeding. By letting the cells sit on ice, the activity of the extract is increased two-fold. The length of time on ice can range from 0 hours to 1 week. The cells were resuspended in 1.5 volumes (packed cell volume) of hypotonic buffer (10 mM HEPES (KOH) pH 7.4; 15 mM KCl; 1.5 mM Mg(OAc)<sub>2</sub>; 0.5 mM Pefabloc (Roche); 2 mM DTT). The cells were allowed to swell for 5 minutes on ice, dounce homogenized with 10 to 100 strokes using a tight-fitting pestle, and spun for 10 minutes at 12000Xg at 4°C in a Sorvall SS-34 rotor. The supernatant was collected with a Pasteur pipet without disturbing the lipid layer, transferred into Eppendorf tubes (50 to 200  $\mu$ l aliquots), and immediately frozen in liquid nitrogen.

15

### 6.1.3. Footprinting

The HeLa translation extract was incubated with the small molecules (at a concentration of 100  $\mu$ M), followed by treatment with chemical modifying agents (dimethyl sulfate [DMS] and kethoxal [KE]). Following chemical modification, rRNA was phenol-chloroform extracted, ethanol precipitated, analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to different regions of the rRNAs and resolved on 6% polyacrylamide gels. The probes used for primer extension cover the entire 18S (7 oligonucleotide primers), 28S (24 oligonucleotide primers), and 5S (one primer) rRNAs are presented in Table 1 (also see, e.g., Gonzalez et al., 1985 Proc Natl Acad Sci U S A. 82(22):7666-70 and McCallum & Maden, 1985, Biochem. J. 232 (3): 725-733). Controls in these experiments include DMSO (a control for changes in rRNA accessibility induced by DMSO), paromomycin (a marker for 18S rRNA binding), and anisomycin (a marker for 28S rRNA binding).

30 Table 1: 18S, 28S, and 5S rRNA primers

5S#1	AAAGCCTACAGCACCC	SEQ ID NO.: 1
28S#1	TACTGAGGGAATCCTGG	SEQ ID NO.: 2
28S#2	TTACCACCCGCTTGGG	SEQ ID NO.: 3
28S#3	GGGGCGGGAAAGATCC	SEQ ID NO.: 4

35

28S#4	CCCCGAGCCACCTTCCC	SEQ ID NO.: 5
28S#5	GGCCCCGGGATTGGCG	SEQ ID NO.: 6
28S#6	CACTGGGGACAGTCCGC	SEQ ID NO.: 7
5 28S#7	CGCGGCGGGCGAGACGGG	SEQ ID NO.: 8
28S#8	GAGGGAAACTCGGAGGG	SEQ ID NO.: 9
28S#9	CATCGGGCGCCTAACCC	SEQ ID NO.: 10
28S#10	CGACGCACACCACACGC	SEQ ID NO.: 11
10 28S#11	CCAAGATCTGCACCTGC	SEQ ID NO.: 12
28S#12	TTACCGCACTGGACGCC	SEQ ID NO.: 13
28S#13	GCCAGAGGCTGTTACC	SEQ ID NO.: 14
28S#14	TGGGGAGGGAGCGAGCGGGCG	SEQ ID NO.: 15
15 28S#15	AAGGGCCCGGCTCGCGTCC	SEQ ID NO.: 16
28S#16	AGGGCGGGGGGACGAACCGC	SEQ ID NO.: 17
28S#17	TTAACACAGTCGGATTCCCCTGG	SEQ ID NO.: 18
28S#18	TTCATCCATTCATGCGCG	SEQ ID NO.: 19
20 28S#19	AGTAGTGGTATTCACCGG	SEQ ID NO.: 20
28S#20	ACGGGAGGTTCTGTCC	SEQ ID NO.: 21
28S#21	ACAATGATAGGAAGAGCCG	SEQ ID NO.: 22
28S#22	AGGCGTTCAGTCATAATCCC	SEQ ID NO.: 23
25 28S#23	TCCGCACCGGACCCCGGTCC	SEQ ID NO.: 24
28S#24	GGGCTAGTTGATTGGCAGGTGAGTTG	SEQ ID NO.: 25
18S#1	TCTCCGGAATCGAACCCCT	SEQ ID NO.: 26
18S#2	ATT ACC GCGGCTGCTGGC	SEQ ID NO.: 27
30 18S#3	TTGGCAAATGCTTCGC	SEQ ID NO.: 28
18S#4	CCGTCAATTCTTTAAGTTTC	SEQ ID NO.: 29
18S#5	AGGGCATCACAGACCTGTTAT	SEQ ID NO.: 30
18S#6	CGACGGGCGGTGTGTAC	SEQ ID NO.: 31
35 18S#7	CCGCAGGTTCACCTACGG	SEQ ID NO.: 32

)

### 6.2. Results

The results of these foot-printing experiments (see, e.g., Figures 2 to 6) indicated that the small molecules involved in modulating translation termination alter the accessibility of the chemical modifying agents to specific nucleotides in the 28S rRNA. 5 More specifically, the regions protected by the small molecules include a conserved region in the vicinity of the peptidyl transferase center (domain V, see, e.g., Figures 2 and 3) implicated in peptide bond formation and a conserved region in domain II (see, e.g., Figures 4, 5, and 6) that may interact with the peptidyl transferase center based on binding of 10 vernamycin B to both these areas (Vannuffel et al., 1994, Nucleic Acids Res. 22(21):4449-53).

### 9. EXAMPLE: HUMAN DISEASE GENES SORTED BY CHROMOSOME

15 Table 2: Genes, Locations and Genetic Disorders on Chromosome 1

Gene	GDB Accession ID	OMIM Link
ABCA4	GDB:370748	MACULAR DEGENERATION, SENILE STARGARDT DISEASE 1; STGD1 ATP BINDING CASSETTE TRANSPORTER; ABCR RETINITIS PIGMENTOSA-19; RP19
ABCD3	GDB:131485	PEROXISOMAL MEMBRANE PROTEIN 1; PXMP1
ACADM	GDB:118958	ACYL-CoA DEHYDROGENASE, MEDIUM-CHAIN; ACADM
AGL	GDB:132644	GLYCOGEN STORAGE DISEASE III
AGT	GDB:118750	ANGIOTENSIN I; AGT
ALDH4A1	GDB:9958827	HYPERPROLINEMIA, TYPE II
ALPL	GDB:118730	PHOSPHATASE, LIVER ALKALINE; ALPL HYPOPHOSPHATASIA, INFANTILE
AMPD1	GDB:119677	ADENOSINE MONOPHOSPHATE DEAMINASE-1; AMPD1
APOA2	GDB:119685	APOLIPOPROTEIN A-II; APOA2
AVSD1	GDB:265302	ATRIOVENTRICULAR SEPTAL DEFECT; AVSD

Gene	GDB Accession ID	OMIM Link
5 BRCD2	GDB:9955322	BREAST CANCER, DUCTAL, 2; BRCD2
C1QA	GDB:119042	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, ALPHA POLYPEPTIDE; C1QA
10 C1QB	GDB:119043	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, BETA POLYPEPTIDE; C1QB
C1QG	GDB:128132	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, GAMMA POLYPEPTIDE; C1QG
15 C8A	GDB:119735	COMPLEMENT COMPONENT-8, DEFICIENCY OF
C8B	GDB:119736	COMPLEMENT COMPONENT-8, DEFICIENCY OF, TYPE II
20 CACNA1S	GDB:126431	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, L TYPE, ALPHA 1S SUBUNIT; CACNA1S PERIODIC PARALYSIS I; MALIGNANT HYPERHERMIA SUSCEPTIBILITY-5; MHS5
CCV	GDB:1336655	CATARACT, CONGENITAL, VOLKMANN TYPE; CCV
25 CD3Z	GDB:119766	CD3Z ANTIGEN, ZETA POLYPEPTIDE; CD3Z
CDC2L1	GDB:127827	PROTEIN KINASE p58; PK58
CHML	GDB:135222	CHOROIDEREMIA-LIKE; CHML
CHS1	GDB:4568202	CHEDIAK-HIGASHI SYNDROME; CHS1
30 CIAS1	GDB:9957338	COLD HYPERSENSITIVITY URTICARIA, DEAFNESS, AND AMYLOIDOSIS
CLCNKB	GDB:698472	CHLORIDE CHANNEL, KIDNEY, B; CLCNKB
35 CMD1A	GDB:434478	CARDIOMYOPATHY, DILATED 1A; CMD1A

35

	Gene	GDB Accession ID	OMIM Link	
5	CMH2	GDB:137324	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 2; CMH2	
	CMM	GDB:119059	MELANOMA, MALIGNANT	
	COL11A1	GDB:120595	COLLAGEN, TYPE XI, ALPHA-1; COL11A1	
10	COL9A2	GDB:138310	COLLAGEN, TYPE IX, ALPHA-2 CHAIN; COL9A2 EPIPHYSEAL DYSPLASIA, MULTIPLE, 2; EDM2	
	CPT2	GDB:127272	MYOPATHY WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE II HYPOGLYCEMIA, HYPOKETOTIC, WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE CARNITINE PALMITOYLTRANSFERASE II; CPT2	
	CRB1	GDB:333930	RETINITIS PIGMENTOSA-12; RP12	
15	CSE	GDB:596182	CHOREOATHETOSIS/SPASTICITY, EPISODIC; CSE	
	CSF3R	GDB:126430	COLONY STIMULATING FACTOR 3 RECEPTOR, GRANULOCYTE; CSF3R	
	CTPA	GDB:9863168	CATARACT, POSTERIOR POLAR	
20	CTSK	GDB:453910	PYCNODYSOSTOSIS CATHEPSIN K; CTSK	
	DBT	GDB:118784	MAPLE SYRUP URINE DISEASE, TYPE 2	
	DIO1	GDB:136449	THYROXINE DEIODINASE TYPE I; TXDI1	
25	DISC1	GDB:9992707	DISORDER-2; SCZD2	
	DPYD	GDB:364102	DIHYDROPYRIMIDINE DEHYDROGENASE; DPYD	
30	EKV	GDB:119106	ERYTHROKERATODERMIA VARIABILIS; EKV	

Gene	GDB Accession ID	OMIM Link
ENO1	GDB:119871	PHOSPHOPYRUVATE HYDRATASE; PPH
5	ENO1P	PHOSPHOPYRUVATE HYDRATASE; PPH
	EPB41	ERYTHROCYTE MEMBRANE PROTEIN BAND 4.1; EPB41 HEREDITARY HEMOLYTIC
	EPHX1	EPOXIDE HYDROLASE 1, MICROSOMAL; EPHX1
10	F13B	FACTOR XIII, B SUBUNIT; F13B
	F5	FACTOR V DEFICIENCY
15	FCGR2A	Fc FRAGMENT OF IgG, LOW AFFINITY IIa, RECEPTOR FOR; FCGR2A
	FCGR2B	Fc FRAGMENT OF IgG, LOW AFFINITY IIa, RECEPTOR FOR; FCGR2A
	FCGR3A	Fc FRAGMENT OF IgG, LOW AFFINITY IIIa, RECEPTOR FOR; FCGR3A
20	FCHL	HYPERLIPIDEMIA, COMBINED
	FH	FUMARATE HYDRATASE; FH LEIOMYOMATA, HEREDITARY MULTIPLE, OF SKIN
25	FMO3	FLAVIN-CONTAINING MONOOXYGENASE 3; FMO3 TRIMETHYLAMINURIA
	FMO4	FLAVIN-CONTAINING MONOOXYGENASE 2; FMO2
	FUCA1	FUCOSIDOSIS
30	FY	BLOOD GROUP--DUFFY SYSTEM; Fy
	GALE	GALACTOSE EPIMERASE DEFICIENCY
	GBA	GAUCHER DISEASE, TYPE I; GD I
35		

Gene	GDB Accession ID	OMIM Link
GFND	GDB:9958222	GLOMERULAR NEPHRITIS, FAMILIAL, WITH FIBRONECTIN DEPOSITS
GJA8	GDB:696369	CATARACT, ZONULAR PULVERULENT 1; CZP1 GAP JUNCTION PROTEIN, ALPHA-8, 50-KD; GJA8
GJB3	GDB:127820	ERYTHROKERATODERMIA VARIABILIS; EKV DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 2; DFNA2
GLC3B	GDB:3801939	GLAUCOMA 3, PRIMARY INFANTILE, B; GLC3B
HF1	GDB:120041	H FACTOR 1; HF1
HMGCL	GDB:138445	HYDROXYMETHYLGLUTARICACIDURIA; HMGCL
HPC1	GDB:5215209	PROSTATE CANCER; PRCA1 PROSTATE CANCER, HEREDITARY 1
HRD	GDB:9862254	HYPOPARATHYROIDISM WITH SHORT STATURE, MENTAL RETARDATION, AND SEIZURES
HRPT2	GDB:125253	HYPERPARATHYROIDISM, FAMILIAL PRIMARY, WITH MULTIPLE OSSIFYING JAW
HSD3B2	GDB:134044	ADRENAL HYPERPLASIA II
HSPG2	GDB:126372	HEPARAN SULFATE PROTEOGLYCAN OF BASEMENT MEMBRANE; HSPG2 MYOTONIC MYOPATHY, DWARFISM, CHONDRODYSTROPHY, AND OCULAR AND FACIAL
KCNQ4	GDB:439046	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 2; DFNA2

Gene	GDB Accession ID	OMIM Link
KCS	GDB:9848740	KENNY-CAFFEY SYNDROME, RECESSIVE FORM
5 KIF1B	GDB:128645	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, A; CMT2A
LAMB3	GDB:251820	LAMININ, BETA 3; LAMB3
10 LAMC2	GDB:136225	LAMININ, GAMMA 2; LAMC2 EPIDERMOLYSIS BULLOSA LETALIS
LGMD1B	GDB:231606	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 1B; LGMD1B
15 LMNA	GDB:132146	LAMIN A/C; LMNA LIPODYSTROPHY, FAMILIAL PARTIAL, DUNNIGAN TYPE; LDP1
LOR	GDB:132049	LORICRIN; LOR
MCKD1	GDB:9859381	POLYCYSTIC KIDNEYS, MEDULLARY TYPE
MCL1	GDB:139137	MYELOID CELL LEUKEMIA 1; MCL1
20 MPZ	GDB:125266	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS MYELIN PROTEIN ZERO; MPZ
MTHFR	GDB:370882	5,10-@METHYLENETETRAHYDROFOLATE REDUCTASE; MTHFR
25 MTR	GDB:119440	METHYLtetrahydrofolate:L-HOMOCYSTEINE S-METHYLTRANSFERASE; MTR
MUTYH	GDB:9315115	ADENOMATOUS POLYPOSIS OF THE COLON; APC
30 MYOC	GDB:5584221	GLAUCOMA 1, OPEN ANGLE; GLC1A MYOCILIN; MYOC
NB	GDB:9958705	NEUROBLASTOMA; NB
35 NCF2	GDB:120223	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-POSITIVE FORM

Gene	GDB Accession ID	OMIM Link
NEM1	GDB:127387	NEMALINE MYOPATHY 1, AUTOSOMAL DOMINANT; NEM1
NPHS2	GDB:9955617	ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 2; ARVD2
<hr/>		
NPPA	GDB:118727	NATRIURETIC PEPTIDE PRECURSOR A; NPPA
NRAS	GDB:119457	ONCOGENE NRAS; NRAS; NRAS1
NTRK1	GDB:127897	ONCOGENE TRK NEUROTROPHIC TYROSINE KINASE, RECEPTOR, TYPE 1; NTRK1 NEUROPATHY, CONGENITAL SENSORY, WITH ANHIDROSIS
OPTA2	GDB:9955577	OSTEOPETROSIS, AUTOSOMAL DOMINANT, TYPE II; OPA2
PBX1	GDB:125351	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1; PBX1
PCHC	GDB:9955586	PHEOCHROMOCYTOMA
PGD	GDB:119486	6-@PHOSPHOGLUCONATE DEHYDROGENASE, ERYTHROCYTE
PHA2A	GDB:9955628	PSEUDOHYPOALDOSTERONISM, TYPE II; PHA2
PHGDH	GDB:9958261	3-@PHOSPHOGLYCERATE DEHYDROGENASE DEFICIENCY
PKLR	GDB:120294	PYRUVATE KINASE DEFICIENCY OF ERYTHROCYTE
PKP1	GDB:4249598	PLAKOPHILIN 1; PKP1
PLA2G2A	GDB:120296	PHOSPHOLIPASE A2, GROUP IIA; PLA2G2A
PLOD	GDB:127821	PROCOLLAGEN-LYSINE, 2-OXOGLUTARATE 5-DIOXYGENASE; PLOD EHRLERS-DANLOS SYNDROME, TYPE VI; E-D VI; EDS VI
PPOX	GDB:118852	PROTOPORPHYRINOGEN OXIDASE; PPOX

Gene	GDB Accession ID	OMIM Link
PPT	GDB:125227	CEROID-LIPOFUSCINOSIS, NEURONAL 1, INFANTILE; CLN1 PALMITOYL-PROTEIN THIOESTERASE; PPT
PRCC	GDB:3888215	PAPILLARY RENAL CELL CARCINOMA; PRCC
PRG4	GDB:9955719	ARTHROPATHY-CAMPTODACTYLY SYNDROME
PSEN2	GDB:633044	ALZHEIMER DISEASE, FAMILIAL, TYPE 4; AD4
PTOS1	GDB:6279920	PTOSIS, HEREDITARY CONGENITAL 1; PTOS1
REN	GDB:120345	RENIN; REN
RFX5	GDB:6288464	REGULATORY FACTOR 5; RFX5
RHD	GDB:119551	RHESUS BLOOD GROUP, D ANTIGEN; RHD
RMD1	GDB:448902	RIPPLING MUSCLE DISEASE-1; RMD1
RPE65	GDB:226519	RETINAL PIGMENT EPITHELIUM-SPECIFIC PROTEIN, 65-KD; RPE65 AMAUROSIS CONGENITA OF LEBER II
SCCD	GDB:9955558	CORNEAL DYSTROPHY, CRYSTALLINE, OF SCHNYDER
SERPINC1	GDB:119024	ANTITHROMBIN III DEFICIENCY
SJS1	GDB:1381631	MYOTONIC MYOPATHY, DWARFISM, CHONDRODYSTROPHY, AND OCULAR AND FACIAL
SLC19A2	GDB:9837779	THIAMINE-RESPONSIVE MEGALOBLASTIC ANEMIA SYNDROME
SLC2A1	GDB:120627	SOLUTE CARRIER FAMILY 2, MEMBER 1; SLC2A1

Gene	GDB Accession ID	OMIM Link
SPTA1	GDB:119601	ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE; HEREDITARY HEMOLYTIC SPECTRIN, ALPHA, ERYTHROCYTIC 1; SPTA1
TAL1	GDB:120759	T-CELL ACUTE LYMPHOCYTIC LEUKEMIA 1; TAL1
TNFSF6	GDB:422178	APOPTOSIS ANTIGEN LIGAND 1; APT1LG1
TNNT2	GDB:221879	TROPONIN-T2, CARDIAC; TNNT2
TPM3	GDB:127872	ONCOGENE TRK TROPOMYOSIN 3; TPM3
TSHB	GDB:120467	THYROID-STIMULATING HORMONE, BETA CHAIN; TSHB
UMPK	GDB:120481	URIDINE MONOPHOSPHATE KINASE; UMPK
UOX	GDB:127539	URATE OXIDASE; UOX
UROD	GDB:119628	PORPHYRIA CUTANEA Tarda; PCT
USH2A	GDB:120483	USHER SYNDROME, TYPE II; USH2
VMGLOM	GDB:9958134	GLOMUS TUMORS, MULTIPLE
VWS	GDB:120532	CLEFT LIP AND/OR PALATE WITH MUCOUS CYSTS OF LOWER LIP
WS2B	GDB:407579	WAARDENBURG SYNDROME, TYPE 2B; WS2B

Table 3: Genes, Locations and Genetic Disorders on Chromosome 2

Gene	GDB Accession ID	Location	OMIM Link
ABCB11	GDB:9864786	2q24-2q24 2q24.3-2q24.3	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATIC 2; PFIC2
ABCG5	GDB:10450298	2p21-2p21	PHYTOSTEROLEMIA
ABCG8	GDB:10450300	2p21-2p21	PHYTOSTEROLEMIA
ACADL	GDB:118745	2q34-2q35	ACYL-CoA DEHYDROGENASE, LONG-CHAIN, DEFICIENCY OF

	Gene	GDB Accession ID	Location	OMIM Link
5	ACP1	GDB:118962	2p25-2p25	PHOSPHATASE, ACID, OF ERYTHROCYTE; ACP1
	AGXT	GDB:127113	2q37.3-2q37.3	OXALOSIS I
	AHHR	GDB:118984	2pter-2q31	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 1; CYP1A1
10	ALMS1	GDB:9865539	2p13-2p12 2p14-2p13 2p13.1-2p13.1	ALSTROM SYNDROME
	ALPP	GDB:119672	2q37.1-2q37.1	ALKALINE PHOSPHATASE, PLACENTAL; ALPP
15	ALS2	GDB:135696	2q33-2q35	AMYOTROPHIC LATERAL SCLEROSIS 2, JUVENILE; ALS2
	APOB	GDB:119686	2p24-2p23 2p24-2p24	APOLIPOPROTEIN B; APOB
	BDE	GDB:9955730	2q37-2q37	BRACHYDACTYLY, TYPE E; BDE
20	BDMR	GDB:533064	2q37-2q37	BRACHYDACTYLY-MENTAL RETARDATION SYNDROME; BDMR
	BJS	GDB:9955717	2q34-2q36	TORTI AND NERVE DEAFNESS
25	BMPR2	GDB:642243	2q33-2q33 2q33-2q34	PULMONARY HYPERTENSION, PRIMARY; PPH1 BONE MORPHOGENETIC RECEPTOR TYPE II; BMPR2
	CHRNA1	GDB:120586	2q24-2q32	CHOLINERGIC RECEPTOR, NICOTINIC, ALPHA POLYPEPTIDE 1; CHRNA1
30	CMCWTD	GDB:11498919	2p22.3-2p21	FAMILIAL CHRONIC MUCOCUTANEOUS, DOMINANT TYPE
	CNGA3	GDB:434398	2q11.2-2q11.2	COLORBLINDNESS, TOTAL CYCLIC NUCLEOTIDE GATED CHANNEL, OLFACTORY, 3; CNG3

	Gene	GDB Accession ID	Location	OMIM Link
5	COL3A1	GDB:118729	2q31-2q32.3 2q32.2-2q32.2	COLLAGEN, TYPE III; COL3A1 EHLERS-DANLOS SYNDROME, TYPE IV, AUTOSOMAL DOMINANT
	COL4A3	GDB:128351	2q36-2q37	COLLAGEN, TYPE IV, ALPHA-3 CHAIN; COL4A3
	COL4A4	GDB:132673	2q35-2q37	COLLAGEN, TYPE IV, ALPHA-4 CHAIN; COL4A4
10	COL6A3	GDB:119066	2q37.3-2q37.3	COLLAGEN, TYPE VI, ALPHA-3 CHAIN; COL6A3 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
	CPS1	GDB:119799	2q33-2q36 2q34-2q35 2q35-2q35	HYPERAMMONEMIA DUE TO CARBAMOYLPHOSPHATE SYNTETASE I DEFICIENCY
15	CRYGA	GDB:119076	2q33-2q35	CRYSTALLIN, GAMMA A; CRYGA
	CRYGEP1	GDB:119808	2q33-2q35	CRYSTALLIN, GAMMA A; CRYGA
	CYP1B1	GDB:353515	2p21-2p21 2p22-2p21 2pter-2qter	GLAUCOMA 3, PRIMARY INFANTILE, A; GLC3A CYTOCHROME P450, SUBFAMILY I (DIOXIN-INDUCIBLE), POLYPEPTIDE 1; CYP1B1
20	CYP27A1	GDB:128129	2q33-2qter	CEREBROTENDINOUS XANTHOMATOSIS
	DBI	GDB:119837	2q12-2q21	DIAZEPAM BINDING INHIBITOR; DBI
25	DES	GDB:119841	2q35-2q35	DESMIN; DES
	DYSF	GDB:340831	2p-2p 2p13-2p13 2pter-2p12	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2B; LGMD2B MUSCULAR DYSTROPHY, LATE-ONSET DISTAL
	EDAR	GDB:9837372	2q11-2q13	DYSPLASIA, HYPOHIDROTIC ECTODERMAL DYSPLASIA, ANHIDROTIC
30				
35				

Gene	GDB Accession ID	Location	OMIM Link
EFEMP1	GDB:1220111	2p16-2p16	DOYNE HONEYCOMB DEGENERATION OF RETINA FIBRILLIN-LIKE; FBNL
5 EIF2AK3	GDB:9956743	2p12-2p12	EPIPHYSEAL DYSPLASIA, MULTIPLE, WITH EARLY-ONSET DIABETES MELLITUS
10 ERCC3	GDB:119881	2q21-2q21	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 3; ERCC3
15 FSHR	GDB:127510	2p21-2p16	FOLLICLE-STIMULATING HORMONE RECEPTOR; FSHR GONADAL DYSGENESIS, XX TYPE
20 GAD1	GDB:119244	2q31-2q31	PYRIDOXINE DEPENDENCY WITH SEIZURES
25 GINGF	GDB:9848875	2p21-2p21	GINGIVAL SON OF SEVENLESS (DROSOPHILA) HOMOLOG 1; SOS1
30 GLC1B	GDB:1297553	2q1-2q13	GLAUCOMA 1, OPEN ANGLE, B; GLC1B
GPD2	GDB:354558	2q24.1-2q24.1	GLYCEROL-3-PHOSPHATE DEHYDROGENASE-2; GPD2
GYPC	GDB:120027	2q14-2q21	BLOOD GROUP--GERBICH; Ge
35 HADHA	GDB:434026	2p23-2p23	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
HADHB	GDB:344953	2p23-2p23	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
HOXD13	GDB:127225	2q31-2q31	HOMEOPBOX-D13; HOXD13 SYNDACTYLY, TYPE II
HPE2	GDB:136066	2p21-2p21	MIDLINE CLEFT SYNDROME

	<b>Gene</b>	<b>GDB Accession ID</b>	<b>Location</b>	<b>OMIM Link</b>
	IGKC	GDB:120088	2p12-2p12 2p11.2-2p11.2	IMMUNOGLOBULIN KAPPA CONSTANT REGION; IGKC
5	IHH	GDB:511203	2q33-2q35 2q35-2q35 2pter-2qter	BRACHYDACTYLY, TYPE A1; BDA1 INDIAN HEDGEHOG, DROSOPHILA, HOMOLOG OF; IHH
	IRS1	GDB:133974	2q36-2q36	INSULIN RECEPTOR SUBSTRATE 1; IRS1
10	ITGA6	GDB:128027	2pter-2qter	INTEGRIN, ALPHA-6; ITGA6
	KHK	GDB:391903	2p23.3-2p23.2	FRUCTOSURIA
	KYNU	GDB:9957925	2q22.2-2q23.3	
	LCT	GDB:120140	2q21-2q21	DISACCHARIDE INTOLERANCE II
15	LHCGR	GDB:125260	2p21-2p21	LUTEINIZING HORMONE/CHORIOGONADOTROPIN RECEPTOR; LHCGR
	LSFC	GDB:9956219	2-2 2p16-2p16	CYTOCHROME c OXIDASE DEFICIENCY, FRENCH-CANADIAN TYPE
20	MSH2	GDB:203983	2p16-2p16 2p22-2p21	COLON CANCER, FAMILIAL, NONPOLYPOSIS TYPE 1; FCC1
	MSH6	GDB:632803	2p16-2p16	G/T MISMATCH-BINDING PROTEIN; GTBP
25	NEB	GDB:120224	2q24.1-2q24.2	NEBULIN; NEB NEMALINE MYOPATHY 2, AUTOSOMAL RECESSIVE; NEM2
	NMTC	GDB:11498336	2q21-2q21	THYROID CARCINOMA, PAPILLARY
	NPHP1	GDB:128050	2q13-2q13	NEPHRONOPHTHISIS, FAMILIAL JUVENILE 1; NPHP1
30	PAFAH1P1	GDB:435099	2p11.2-2p11.2	PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
	PAX3	GDB:120495	2q36-2q36 2q35-2q35	KLEIN-WAARDENBURG SYNDROME WAARDENBURG SYNDROME; WS1
35				

	<b>Gene</b>	<b>GDB Accession ID</b>	<b>Location</b>	<b>OMIM Link</b>
5	PAX8	GDB:136447	2q12-2q14	PAIRED BOX HOMEOTIC GENE 8; PAX8
	PMS1	GDB:386403	2q31-2q33	POSTMEIOTIC SEGREGATION INCREASED (S. CERESIAE)-1; PMS1
	PNKD	GDB:5583973	2q33-2q35	CHOREOATHETOSIS, FAMILIAL PAROXYSMAL; FPD1
10	PPH1	GDB:1381541	2q31-2q32 2q33-2q33	PULMONARY HYPERTENSION, PRIMARY; PPH1
	PROC	GDB:120317	2q13-2q21 2q13-2q14	PROTEIN C DEFICIENCY, CONGENITAL THROMBOTIC DISEASE DUE TO
15	REG1A	GDB:132455	2p12-2p12	REGENERATING ISLET-DERIVED 1-ALPHA; REG1A
	SAG	GDB:120365	2q37.1-2q37.1	S-ANTIGEN; SAG
20	SFTPB	GDB:120374	2p12-2p11.2	SURFACTANT-ASSOCIATED PROTEIN, PULMONARY-3; SFTP3
	SLC11A1	GDB:371444	2q35-2q35	CIRRHOSIS, PRIMARY; PBC NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 1; NRAMP1
25	SLC3A1	GDB:202968	2p16.3-2p16.3 2p21-2p21	SOLUTE CARRIER FAMILY 3, MEMBER 1; SLC3A1 CYSTINURIA; CSNU
	SOS1	GDB:230004	2p22-2p21	GINGIVAL SON OF SEVENLESS (DROSOPHILA) HOMOLOG 1; SOS1
30	SPG4	GDB:230127	2p24-2p21	SPASTIC PARAPLEGIA-4, AUTOSOMAL DOMINANT; SPG4
	SRD5A2	GDB:127343	2p23-2p23	PSEUDOVAGINAL PERINEOSCROTAL HYPOSPADIAS; PPSH

Gene	GDB Accession ID	Location	OMIM Link
TCL4	GDB:136378	2q34-2q34	T-CELL LEUKEMIA/LYMPHOMA-4; TCL4
TGFA	GDB:120435	2p13-2p13	TRANSFORMING GROWTH FACTOR, ALPHA; TGFA
TMD	GDB:9837196	2q31-2q31	TIBIAL MUSCULAR DYSTROPHY, TARDIVE
TPO	GDB:120446	2p25-2p25 2p25-2p24	THYROID HORMONOGENESIS, GENETIC DEFECT IN, IIA
UGT1	GDB:120007	2q37-2q37	UDP GLUCURONOSYLTRANSFERASE 1 FAMILY, A1; UGT1A1
UV24	GDB:9955737	2pter-2qter	UV-DAMAGE, EXCISION REPAIR OF, UV-24
WSS	GDB:9955707	2q32-2q32	WRINKLY SKIN SYNDROME; WSS
XDH	GDB:266386	2p23-2p22	XANTHINURIA
ZAP70	GDB:433738	2q11-2q13 2q12-2q12	SYK-RELATED TYROSINE KINASE; SRK
ZFHX1B	GDB:9958310	2q22-2q22	DISEASE, MICROCEPHALY, AND IRIS COLOBOMA

Table 4: Genes, Locations and Genetic Disorders on Chromosome 3

Gene	GDB Accession ID	Location	OMIM Link
ACAA1	GDB:119643	3p23-3p22	PEROXISOMAL 3-OXOACYL-COENZYME A THIOLASE DEFICIENCY
AGTR1	GDB:132359	3q21-3q25	ANGIOTENSIN II RECEPTOR, VASCULAR TYPE 1; AT2R1
AHSG	GDB:118985	3q27-3q27	ALPHA-2-HS-GLYCOPROTEIN; AHSG
AMT	GDB:132138	3p21.3-3p21.2 3p21.2-3p21.1	HYPERGLYCEMIA, ISOLATED NONKETOTIC, TYPE II; NKH2
ARP	GDB:9959049	3p21.1-3p21.1	ARGININE-RICH PROTEIN
BBS3	GDB:376501	3p-3p 3p12.3-3q11.1	BARDET-BIEDL SYNDROME, TYPE 3; BBS3

	Gene	GDB Accession ID	Location	OMIM Link
	BCHE	GDB:120558	3q26.1-3q26.2	BUTYRYLCHOLINESTERASE; BCHE
5	BCPM	GDB:433809	3q21-3q21	BENIGN CHRONIC PEMPHIGUS; BCPM
	BTD	GDB:309078	3p25-3p25	BIOTINIDASE; BTD
10	CASR	GDB:134196	3q21-3q24	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL; HHC1
	CCR2	GDB:337364	3p21-3p21	CHEMOKINE (C-C) RECEPTOR 2; CMKBR2
15	CCR5	GDB:1230510	3p21-3p21	CHEMOKINE (C-C) RECEPTOR 5; CMKBR5
	CDL1	GDB:136344	3q26.3-3q26.3	DE LANGE SYNDROME; CDL
20	CMT2B	GDB:604021	3q13-3q22	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, B; CMT2B
	COL7A1	GDB:128750	3p21-3p21 3p21.3-3p21.3	COLLAGEN, TYPE VII, ALPHA-1; COL7A1
25	CP	GDB:119069	3q23-3q25 3q21-3q24	CERULOPLASMIN; CP
	CRV	GDB:11498333	3p21.3-3p21.1	VASCULOPATHY, RETINAL, WITH CEREBRAL LEUKODYSTROPHY
30	CTNNB1	GDB:141922	3p22-3p22 3p21.3-3p21.3	CATENIN, BETA 1; CTNNB1
	DEM	GDB:681157	3p12-3q11	DEMENTIA, FAMILIAL NONSPECIFIC; DEM
35	ETM1	GDB:9732523	3q13-3q13	TREMOR, HEREDITARY ESSENTIAL 1; ETM1
	FANCD2	GDB:698345	3p25.3-3p25.3 3pter-3p24.2	FANCONI PANCYTOPENIA, COMPLEMENTATION GROUP D
	FIH	GDB:9955790	3q13-3q13	HYPOPARATHYROIDISM, FAMILIAL ISOLATED; FIH
	FOXL2	GDB:129025	3q23-3q23 3q22-3q23	BLEPHAROPHIMOSIS, EPICANTHUS INVERSUS, AND PTOSIS; BPES

Gene	GDB Accession ID	Location	OMIM Link
GBE1	GDB:138442	3p12-3p12	GLYCOGEN STORAGE DISEASE IV
5 GLB1	GDB:119987	3p22-3p21.33 3p21.33-3p21.33	GANGLIOSIDOSIS, GENERALIZED GM1, TYPE I
GLC1C	GDB:3801941	3q21-3q24	GLAUCOMA 1, OPEN ANGLE, C; GLC1C
10 GNAI2	GDB:120516	3p21.3-3p21.2	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-INHIBITING, POLYPEPTIDE-2;
15 GNAT1	GDB:119277	3p21.3-3p21.2	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-TRANSDUCING, POLYPEPTIDE
15 GP9	GDB:126370	3pter-3qter	PLATELET GLYCOPROTEIN IX; GP9
15 GPX1	GDB:119282	3q11-3q12 3p21.3-3p21.3	GLUTATHIONE PEROXIDASE; GPX1
20 HGD	GDB:203935	3q21-3q23	ALKAPTONURIA; AKU
20 HRG	GDB:120055	3q27-3q27	HISTIDINE-RICH GLYCOPROTEIN; HRG; HRGP
25 ITIH1	GDB:120107	3p21.2-3p21.1	INTER-ALPHA-TRYPSIN INHIBITOR, HEAVY CHAIN-1; ITIH1; IATIH; ITIH
25 KNG	GDB:125256	3q27-3q27	FLAUJEAC FACTOR DEFICIENCY
30 LPP	GDB:1391795	3q27-3q28	LIM DOMAIN-CONTAINING PREFERRED TRANSLOCATION PARTNER IN LIPOMA; LPP
30 LRS1	GDB:682448	3p21.1-3p14.1	LARSEN SYNDROME, AUTOSOMAL DOMINANT; LRS1
MCCC1	GDB:135989	3q27-3q27 3q25-3q27	BETA-METHYLCROTONYLGLYC INURIA I
35 MDS1	GDB:250411	3q26-3q26	MYELODYSPLASIA SYNDROME 1; MDS1

Gene	GDB Accession ID	Location	OMIM Link
MHS4	GDB:574245	3q13.1-3q13.1	HYPERTHERMIA SUSCEPTIBILITY-4; MHS4
5 MITF	GDB:214776	3p14.1-3p12	MICROPHTHALMIA-ASSOCIATED TRANSCRIPTION FACTOR; MITF WAARDENBURG SYNDROME, TYPE II; WS2
MLH1	GDB:249617	3p23-3p22 3p21.3-3p21.3	COLON CANCER, FAMILIAL, NONPOLYPOSIS TYPE 2; FCC2
10 MYL3	GDB:120218	3p21.3-3p21.2	MYOSIN, LIGHT CHAIN, ALKALI, VENTRICULAR AND SKELETAL SLOW; MYL3
MYMY	GDB:11500610	3p26-3p24.2	DISEASE
OPA1	GDB:118848	3q28-3q29	OPTIC ATROPHY 1; OPA1
15 PBXP1	GDB:125352	3q22-3q23	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1; PBX1
PCCB	GDB:119474	3q21-3q22	GLYCINEMIA, KETOTIC, II
POU1F1	GDB:129070	3p11-3p11	POU DOMAIN, CLASS 1, TRANSCRIPTION FACTOR 1; POU1F1
20 PPARG	GDB:1223810	3p25-3p25	CANCER OF COLON PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR, GAMMA; PPARG
PROS1	GDB:120721	3p11-3q11 3p11.1-3q11.2	PROTEIN S, ALPHA; PROS1
25 PTHR1	GDB:138128	3p22-3p21.1	METAPHYSEAL CHONDRODYSPLASIA, MURK JANSEN TYPE PARATHYROID HORMONE RECEPTOR 1; PTHR1
RCA1	GDB:230233	3p14.2-3p14.2	RENAL CARCINOMA, FAMILIAL, ASSOCIATED 1; RCA1
30 RHO	GDB:120347	3q21.3-3q24	RHODOPSIN; RHO
SCA7	GDB:454471	3p21.1-3p12	SPINOCEREBELLAR ATAXIA 7; SCA7
35 SCLC1	GDB:9955750	3p23-3p21	SMALL-CELL CANCER OF THE LUNG; SCCL

	Gene	GDB Accession ID	Location	OMIM Link
5	SCN5A	GDB:132152	3p21-3p21	SODIUM CHANNEL, VOLTAGE-GATED, TYPE V, ALPHA POLYPEPTIDE; SCNSA
10	SI	GDB:120377	3q25.2-3q26.2	DISACCHARIDE INTOLERANCE I
15	SLC25A20	GDB:6503297	3p21.31-3p21.31	CARNITINE-ACYLCARNITINE TRANSLOCASE; CACT
20	SLC2A2	GDB:119995	3q26.2-3q27 3q26.1-3q26.3	SOLUTE CARRIER FAMILY 2, MEMBER 2; SLC2A2 FANCONI-BICKEL SYNDROME; FBS
25	TF	GDB:120432	3q21-3q21	TRANSFERRIN; TF
30	TGFBR2	GDB:224909	3p22-3p22 3pter-3p24.2	TRANSFORMING GROWTH FACTOR-BETA RECEPTOR, TYPE II; TGFBR2
35	THPO	GDB:374007	3q26.3-3q27	THROMBOPOIETIN; THPO
	THRB	GDB:120731	3p24.1-3p22 3p24.3-3p24.3	THYROID HORMONE RECEPTOR, BETA; THRB
	TKT	GDB:132402	3p14.3-3p14.3	WERNICKE-KORSAKOFF SYNDROME
	TM4SF1	GDB:250815	3q21-3q25	TUMOR-ASSOCIATED ANTIGEN L6; TAAL6
	TRH	GDB:128072	3pter-3qter	THYROTROPIN-RELEASING HORMONE DEFICIENCY
	UMPS	GDB:120482	3q13-3q13	OROTICACIDURIA I
	UQCRC1	GDB:141850	3p21.3-3p21.2 3p21.3-3p21.3	UBIQUINOL-CYTOCHROME c REDUCTASE CORE PROTEIN I; UQCRC1
	USH3A	GDB:392645	3q21-3q25	USHER SYNDROME, TYPE III; USH3
	VHL	GDB:120488	3p26-3p25	VON HIPPEL-LINDAU SYNDROME; VHL
	WS2A	GDB:128053	3p14.2-3p13	MICROPHTHALMIA-ASSOCIATED TRANSCRIPTION FACTOR; MITF WAARDENBURG SYNDROME, TYPE II; WS2

Gene	GDB Accession ID	Location	OMIM Link
XPC	GDB:134769	3p25.1-3p25.1	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP C; XPC
ZNF35	GDB:120507	3p21-3p21	ZINC FINGER PROTEIN-35; ZNF35

5 Table 5: Genes, Locations and Genetic Disorders on Chromosome 4

Gene	GDB Accession ID	Location	OMIM Link
ADH1B	GDB:119651	4q21-4q23 4q22-4q22	ALCOHOL DEHYDROGENASE-2; ADH2
ADH1C	GDB:119652	4q21-4q23 4q22-4q22	ALCOHOL DEHYDROGENASE-3; ADH3
AFP	GDB:119660	4q11-4q13	ALPHA-FETOPROTEIN; AFP
AGA	GDB:118981	4q23-4q35 4q32-4q33	ASPARTYLGLUCOSAMINURIA ; AGU
AIH2	GDB:118751	4q11-4q13 4q13.3-4q21.2	AMELOGENESIS IMPERFECTA 2, HYPOPLASTIC LOCAL, AUTOSOMAL DOMINANT;
ALB	GDB:118990	4q11-4q13	ALBUMIN; ALB
ASMD	GDB:119705	4q-4q 4q28-4q31	ANTERIOR SEGMENT OCULAR DYSGENESIS; ASOD
BFHD	GDB:11498907	4q34.1-4q35	DYSPLASIA, BEUKES TYPE
CNGA1	GDB:127557	4p14-4q13	CYCLIC NUCLEOTIDE GATED CHANNEL, PHOTORECEPTOR, cGMP GATED, 1; CNCG1
CRBM	GDB:9958132	4p16.3-4p16.3	CHERUBISM
DCK	GDB:126810	4q13.3-4q21.1	DEOXYCYTIDINE KINASE; DCK
DFNA6	GDB:636175	4p16.3-4p16.3	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 6; DFNA6
DSPP	GDB:5560457	4pter-4qter 4q21.3-4q21.3	DENTIN PHOSPHOPROTEIN; DPP DENTINOGENESIS IMPERFECTA; DGI1
DTDP2	GDB:9955810	4q-4q	DENTIN DYSPLASIA, TYPE II
ELONG	GDB:11498700	4q24-4q24	

Gene	GDB Accession ID	Location	OMIM Link
ENAM	GDB:9955259	4q21-4q21	AMELOGENESIS IMPERFECTA 2, HYPOPLASTIC LOCAL, AUTOSOMAL DOMINANT; AMELOGENESIS IMPERFECTA, HYPOPLASTIC TYPE
ETFDH	GDB:135992	4q32-4q35	GLUTARICACIDURIA IIC; GA IIC
EVC	GDB:555573	4p16-4p16	ELLIS-VAN CREVELD SYNDROME; EVC
F11	GDB:119891	4q35-4q35	PTA DEFICIENCY
FABP2	GDB:119127	4q28-4q31	FATTY ACID BINDING PROTEIN 2, INTESTINAL; FABP2
FGA	GDB:119129	4q28-4q28	AMYLOIDOSIS, FAMILIAL; VISCELAR FIBRINOGEN, A ALPHA POLYPEPTIDE; FGA
FGB	GDB:119130	4q28-4q28	FIBRINOGEN, B BETA POLYPEPTIDE; FGB
FGFR3	GDB:127526	4p16.3-4p16.3	ACHONDROPLASIA; ACH BLADDER CANCER; FIBROBLAST GROWTH FACTOR RECEPTOR-3; FGFR3
FGG	GDB:119132	4q28-4q28	FIBRINOGEN, G GAMMA POLYPEPTIDE; FGG
FSHMD1A	GDB:119914	4q35-4q35	FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY 1A; FSHMD1A
GC	GDB:119263	4q12-4q13 4q12-4q12	GROUP-SPECIFIC COMPONENT; GC
GNPTA	GDB:119280	4q21-4q23	MUCOLIPIDOSIS II; ML2; ML II
GNRHR	GDB:136456	4q13-4q13 4q21.2-4q21.2	GONADOTROPIN-RELEASING HORMONE RECEPTOR; GNRHR
GYPA	GDB:118890	4q28-4q31 4q28.2-4q31.1	BLOOD GROUP-MN LOCUS; MN
HCA	GDB:9954675	4q33-4qter	HYPERCALCIURIA, FAMILIAL IDIOPATHIC
HCL2	GDB:119305	4q28-4q31 4q-4q	HAIR COLOR-2; HCL2

Gene	GDB Accession ID	Location	OMIM Link
HD	GDB:119307	4p16.3-4p16.3	HUNTINGTON DISEASE; HD
HTN3	GDB:125601	4q12-4q21	HISTATIN-3; HTN3
HVBS6	GDB:120687	4q32-4q32	HEPATOCELLULAR CARCINOMA-2; HCC2
IDUA	GDB:119327	4p16.3-4p16.3	MUCOPOLYSACCHARIDOSIS TYPE I; MPS I
IF	GDB:120077	4q24-4q25 4q25-4q25	COMPLEMENT COMPONENT-3 INACTIVATOR, DEFICIENCY OF
JPD	GDB:120113	4pter-4qter 4q12-4q13	PERIODONTITIS, JUVENILE; JPD
KIT	GDB:120117	4q12-4q12	V-KIT HARDY-ZUCKERMAN 4 FELINE SARCOMA VIRAL ONCOGENE HOMOLOG; KIT
KLKB1	GDB:127575	4q34-4q35 4q35-4q35	FLETCHER FACTOR DEFICIENCY
LQT4	GDB:682072	4q25-4q27	SYNDROME WITHOUT PSYCHOMOTOR RETARDATION
MANBA	GDB:125261	4q21-4q25	MANNOSIDOSIS, BETA; MANB1
MLLT2	GDB:136792	4q21-4q21	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 2; MLLT2
MSX1	GDB:120683	4p16.3-4p16.1 4p16.1-4p16.1	MSH, DROSOPHILA, HOME BOX, HOMOLOG OF, 1; MSX1
MTP	GDB:228961	4q24-4q24	MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN, 88 KD; MTP
NR3C2	GDB:120188	4q31-4q31 4q31.1-4q31.1	PSEUDOHYPOALDOSTERONISM, TYPE I, AUTOSOMAL RECESSIVE; PHA1
PBT	GDB:120260	4q12-4q21	PIEBALD TRAIT; PBT

Gene	GDB Accession ID	Location	OMIM Link
PDE6B	GDB:125915	4p16.3-4p16.3	NIGHTBLINDNESS, CONGENITAL STATIONARY; CSNB3 PHOSPHODIESTERASE 6B, cGMP-SPECIFIC, ROD, BETA; PDE6B
5 PEE1	GDB:7016765	4q31-4q34 4q25-4qter	1; PEE1
10 PITX2	GDB:134770	4q25-4q27 4q25-4q26 4q25-4q25	IRIDOGONIODYSGENESIS, TYPE 2; IRID2 RIEGER SYNDROME, TYPE 1; RIEG1 RIEG BICOID-RELATED HOMEobox TRANSCRIPTION FACTOR 1; RIEG1 HOMEOP BOX 2
15 PKD2	GDB:118851	4q21-4q23	POLYCYSTIC KIDNEY DISEASE 2; PKD2
QDPR	GDB:120331	4p15.3-4p15.3 4p15.31-4p15.31	PHENYLKETONURIA II
20 SGCB	GDB:702072	4q12-4q12	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2E; LGMD2E
25 SLC25A4	GDB:119680	4q35-4q35	ADENINE NUCLEOTIDE TRANSLOCATOR 1; ANTI PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA; PEO
SNCA	GDB:439047	4q21.3-4q22 4q21-4q21	SYNUCLEIN, ALPHA; SNCA PARKINSON DISEASE, FAMILIAL, TYPE 1; PARK1
25 SOD3	GDB:125291	4p16.3-4q21	SUPEROXIDE DISMUTASE, EXTRACELLULAR; SOD3
STATH	GDB:120391	4q11-4q13	STATHERIN; STATH; STR
30 TAPVR1	GDB:392646	4p13-4q11	ANOMALOUS PULMONARY VENOUS RETURN; APVR
TYS	GDB:119624	4q-4q	SCLEROTYLOSIS; TYS
35 WBS2	GDB:132426	4q33-4q35.1	WILLIAMS-BEUREN SYNDROME; WBS
WFS1	GDB:434294	4p-4p 4p16-4p16	DIABETES MELLITUS AND INSIPIDUS WITH OPTIC ATROPHY AND DEAFNESS

Gene	GDB Accession ID	Location	OMIM Link
WHCR	GDB:125355	4p16.3-4p16.3	WOLF-HIRSCHHORN SYNDROME; WHS

5 Table 6: Genes, Locations and Genetic Disorders on Chromosome 5

	Gene	GDB Accession ID	OMIM Link
10	ADAMTS2	GDB:9957209	EHLERS-DANLOS SYNDROME, TYPE VII, AUTOSOMAL RECESSIVE
15	ADRB2	GDB:120541	BETA-2-ADRENERGIC RECEPTOR; ADRB2
20	AMCN	GDB:9836823	ARTHROGRYPOSIS MULTIPLEX CONGENITA, NEUROGENIC TYPE
25	AP3B1	GDB:9955590	HERMANSKY-PUDLAK SYNDROME; HPS
30	APC	GDB:119682	ADENOMATOUS POLYPOSIS OF THE COLON; APC
35	ARSB	GDB:119008	MUCOPOLYSACCHARIDOSIS TYPE VI; MPS VI
	B4GALT7	GDB:9957653	SYNDROME, PROGEROID FORM
	BHR1	GDB:9956078	ASTHMA
	C6	GDB:119045	COMPLEMENT COMPONENT-6, DEFICIENCY OF
	C7	GDB:119046	COMPLEMENT COMPONENT-7, DEFICIENCY OF
	CCAL2	GDB:5584265	CHONDROCALCINOSIS, FAMILIAL ARTICULAR
	CKN1	GDB:128586	COCKAYNE SYNDROME, TYPE I; CKN1
	CMDJ	GDB:9595425	CRANIOMETAPHYSEAL DYSPLASIA, JACKSON TYPE; CMDJ
	CRHBP	GDB:127438	CORTICOTROPIN RELEASING HORMONE-BINDING PROTEIN; CRHBP
	CSF1R	GDB:120600	COLONY-STIMULATING FACTOR-1 RECEPTOR; CSF1R

35

Gene	GDB Accession ID	OMIM Link
DHFR	GDB:119845	DIHYDROFOLATE REDUCTASE; DHFR
5 DIAPH1	GDB:9835482	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 1; DFNA1 DIAPHANOUS, DROSOPHILA, HOMOLOG OF, 1
10 DTR	GDB:119853	DIPHTHERIA TOXIN SENSITIVITY; DTS
EOS	GDB:9956083	EOSINOPHILIA, FAMILIAL
15 ERVR	GDB:9835857	HYALOIDEORETINAL DEGENERATION OF WAGNER
F12	GDB:119892	HAGEMAN FACTOR DEFICIENCY
15 FBN2	GDB:128122	CONTRACTURAL ARACHNODACTYLY, CONGENITAL; CCA
20 GDNF	GDB:450609	GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR; GDNF
GHR	GDB:119984	GROWTH HORMONE RECEPTOR; GHR
25 GLRA1	GDB:118801	GLYCINE RECEPTOR, ALPHA-1 SUBUNIT; GLRA1 KOK DISEASE
GM2A	GDB:120000	TAY-SACHS DISEASE, AB VARIANT
HEXB	GDB:119308	SANDHOFF DISEASE
25 HSD17B4	GDB:385059	17-@BETA-HYDROXYSTEROID DEHYDROGENASE IV; HSD17B4
ITGA2	GDB:128031	INTEGRIN, ALPHA-2; ITGA2
KFS	GDB:9958987	VERTEBRAL FUSION
30 LGMD1A	GDB:118832	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 1A; LGMD1A
LOX	GDB:119367	LYSYL OXIDASE; LOX
LTC4S	GDB:384080	LEUKOTRIENE C4 SYNTHASE; LTC4S

Gene	GDB Accession ID	OMIM Link
MAN2A1	GDB:136413	MANNOSIDASE, ALPHA, II; MANA2 DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE II
MCC	GDB:128163	MUTATED IN COLORECTAL CANCERS; MCC
MCCC2	GDB:135990	II
MSH3	GDB:641986	MutS, E. COLI, HOMOLOG OF, 3; MSH3
MSX2	GDB:138766	MSH (DROSOPHILA) HOMEO BOX HOMOLOG 2; MSX2 PARIETAL FORAMINA, SYMMETRIC; PFM
NR3C1	GDB:120017	GLUCOCORTICOID RECEPTOR; GRL
PCSK1	GDB:128033	PROPROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 1; PCSK1
PDE6A	GDB:120265	PHOSPHODIESTERASE 6A, cGMP-SPECIFIC, ROD, ALPHA; PDE6A
PFBI	GDB:9956096	INTENSITY OF INFECTION IN
RASA1	GDB:120339	RAS p21 PROTEIN ACTIVATOR 1; RASA1
SCZD1	GDB:120370	DISORDER-1; SCZD1
SDHA	GDB:378037	SUCCINATE DEHYDROGENASE COMPLEX, SUBUNIT A, FLAVOPROTEIN; SDHA
SGCD	GDB:5886421	SARCOGLYCAN, DELTA; SGCD
SLC22A5	GDB:9863277	CARNITINE DEFICIENCY, SYSTEMIC, DUE TO DEFECT IN RENAL REABSORPTION
SLC26A2	GDB:125421	DIASTROPHIC DYSPLASIA; DTD EPIPHYSEAL DYSPLASIA, MULTIPLE; MED NEONATAL OSSEOUS DYSPLASIA I ACHONDROGENESIS, TYPE IB; ACG1B

	Gene	GDB Accession ID	OMIM Link
	SLC6A3	GDB:132445	SOLUTE CARRIER FAMILY 6, MEMBER 3; SLC6A3 DEFICIT-HYPERACTIVITY DISORDER; ADHD
5	SM1	GDB:9834488	SCHISTOSOMA MANSONI SUSCEPTIBILITY/RESISTANCE
	SMA@	GDB:120378	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 1, TELOMERIC; SMN1
10	SMN1	GDB:5215173	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 1, TELOMERIC; SMN1
	SMN2	GDB:5215175	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 2, CENTROMERIC; SMN2
15	SPINK5	GDB:9956114	NETHERTON DISEASE
	TCOF1	GDB:127390	TREACHER COLLINS-FRANCESCHETTI SYNDROME 1; TCOF1
20	TGFBI	GDB:597601	CORNEAL DYSTROPHY, GRANULAR TYPE CORNEAL DYSTROPHY, LATTICE TYPE I; CDL1 TRANSFORMING GROWTH FACTOR, BETA-INDUCED, 68 KD; TGFBI

Table 7: Genes, Locations and Genetic Disorders on Chromosome 6

	Gene	GDB Accession ID	OMIM Link
25	ALDH5A1	GDB:454767	SUCCINIC SEMIALDEHYDE DEHYDROGENASE, NAD(+-)DEPENDENT; SSADH
	ARG1	GDB:119006	ARGININEMIA
30	AS	GDB:135697	ANKYLOSING Spondylitis; AS
	ASSP2	GDB:119017	CITRULLINEMIA
	BCKDHB	GDB:118759	MAPLE SYRUP URINE DISEASE, TYPE IB
35	BF	GDB:119726	GLYCINE-RICH BETA-GLYCOPROTEIN; GBG

Gene	GDB Accession ID	OMIM Link
C2	GDB:119731	COMPLEMENT COMPONENT-2, DEFICIENCY OF
5 C4A	GDB:119732	COMPLEMENT COMPONENT 4A; C4A
CDKN1A	GDB:266550	CYCLIN-DEPENDENT KINASE INHIBITOR 1A; CDKN1A
COL10A1	GDB:128635	COLLAGEN, TYPE X, ALPHA 1; COL10A1
10 COL11A2	GDB:119788	COLLAGEN, TYPE XI, ALPHA-2; COL11A2 STICKLER SYNDROME, TYPE II; STL2 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 13; DFNA13
15 CYP21A2	GDB:120605	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 21-HYDROXYLASE DEFICIENCY
DYX2	GDB:437584	DYSLEXIA, SPECIFIC, 2; DYX2
EJM1	GDB:119864	MYOCLONIC EPILEPSY, JUVENILE; EJM1
20 ELOVL4	GDB:11499609	STARGARDT DISEASE 3; STGD3
EPM2A	GDB:3763331	EPILEPSY, PROGRESSIVE MYOCLONIC 2; EPM2
ESR1	GDB:119120	ESTROGEN RECEPTOR; ESR
25 EYA4	GDB:700062	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 10; DFNA10
F13A1	GDB:120614	FACTOR XIII, A1 SUBUNIT; F13A1
FANCE	GDB:1220236	FANCONI ANEMIA, COMPLEMENTATION GROUP E; FACE
30 GCLC	GDB:132915	GAMMA-GLUTAMYL CYSTEINE SYNTHETASE DEFICIENCY, HEMOLYTIC ANEMIA DUE
GJA1	GDB:125196	GAP JUNCTION PROTEIN, ALPHA-1, 43 KD; GJA1
35 GLYS1	GDB:136421	GLYCOSURIA, RENAL

Gene	GDB Accession ID	OMIM Link
GMPR	GDB:127058	GUANINE MONOPHOSPHATE REDUCTASE
5 GSE	GDB:9956235	DISEASE; CD
HCR	GDB:9993306	PSORIASIS, SUSCEPTIBILITY TO HFE GDB:119309 HEMOCHROMATOSIS; HFE
10 HLA-A	GDB:119310	HLA-A HISTOCOMPATIBILITY TYPE; HLAA HLA-DPB1 GDB:120636 LA-DP HISTOCOMPATIBILITY TYPE, BETA-1 SUBUNIT
HLA-DRA	GDB:120641	HLA-DR HISTOCOMPATIBILITY TYPE; HLA-DRA
15 HPFH	GDB:9849006	HETEROCELLULAR HEREDITARY PERSISTENCE OF FETAL HEMOGLOBIN
ICS1	GDB:136433	IMMOTILE CILIA SYNDROME-1; ICS1
IDDM1	GDB:9953173	DIABETES MELLITUS, JUVENILE-ONSET INSULIN-DEPENDENT; IDDM
20 IFNGR1	GDB:120688	INTERFERON, GAMMA, RECEPTOR-1; IFNGR1
IGAD1	GDB:6929077	SELECTIVE DEFICIENCY OF
IGF2R	GDB:120083	INSULIN-LIKE GROWTH FACTOR 2 RECEPTOR; IGF2R
25 ISCW	GDB:9956158	SUPPRESSION; IS
LAMA2	GDB:132362	LAMININ, ALPHA 2; LAMA2
LAP	GDB:9958992	LARYNGEAL ADDUCTOR PARALYSIS; LAP
30 LCA5	GDB:11498764	AMAUROSIS CONGENITA OF LEBER I
LPA	GDB:120699	APOLIPOPROTEIN(a); LPA
MCDR1	GDB:131406	MACULAR DYSTROPHY, RETINAL, 1, NORTH CAROLINA TYPE; MCDR1
35 MOCS1	GDB:9862235	MOLYBDENUM COFACTOR DEFICIENCY

Gene	GDB Accession ID	OMIM Link
MUT	GDB:120204	METHYLMALONICACIDURIA DUE TO METHYLMALONIC CoA MUTASE DEFICIENCY
5 MYB	GDB:119441	V-MYB AVIAN MYELOBLASTOSIS VIRAL ONCOGENE HOMOLOG; MYB
NEU1	GDB:120230	NEURAMINIDASE DEFICIENCY
10 NKS1	GDB:128100	SUSCEPTIBILITY TO LYSIS BY ALLOREACTIVE NATURAL KILLER CELLS; EC1
NYS2	GDB:9848763	NYSTAGMUS, CONGENITAL
OA3	GDB:136429	ALBINISM, OCULAR, AUTOSOMAL RECESSIVE; OAR
15 ODDD	GDB:6392584	OCULODENTODIGITAL DYSPLASIA; ODDD
OFC1	GDB:120247	OROFACIAL CLEFT 1; OFC1
PARK2	GDB:6802742	PARKINSONISM, JUVENILE
20 PBCA	GDB:9956321	BETA CELL AGENESIS WITH NEONATAL DIABETES MELLITUS
PBCRA1	GDB:3763333	CHORIORETINAL ATROPHY, PROGRESSIVE BIFOCAL; CRAPB
PDB1	GDB:136349	DISEASE OF BONE; PDB
PEX3	GDB:9955507	ZELLWEGER SYNDROME; ZS
25 PEX6	GDB:5592414	ZELLWEGER SYNDROME; ZS PEROXIN-6; PEX6
PEX7	GDB:6155803	RHIZOMELIC CHONDRODYSPLASIA PUNCTATA; RCDP PEROXIN-7; PEX7
PKHD1	GDB:433910	POLYCYSTIC KIDNEY AND HEPATIC DISEASE-1; PKHD1
30 PLA2G7	GDB:9958829	PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, SUBUNIT
PLG	GDB:119498	PLASMINOGEN; PLG
POLH	GDB:6963323	PIGMENTOSUM WITH NORMAL DNA REPAIR RATES

Gene	GDB Accession ID	OMIM Link
PPAC	GDB:9956248	ARTHROPATHY, PROGRESSIVE PSEUDORHEUMATOID, OF CHILDHOOD
5 PSORS1	GDB:6381310	PSORIASIS, SUSCEPTIBILITY TO
PUJO	GDB:9956231	MULTICYSTIC RENAL DYSPLASIA, BILATERAL; MRD
RCD1	GDB:333929	RETINAL CONE DEGENERATION
10 RDS	GDB:118863	RETINAL DEGENERATION, SLOW; RDS
RHAG	GDB:136011	RHESUS BLOOD GROUP-ASSOCIATED GLYCOPROTEIN; RHAG RH-NULL, REGULATOR TYPE; RHN
15 RP14	GDB:433713	RETINITIS PIGMENTOSA-14; RP14 TUBBY-LIKE PROTEIN 1; TULP1
RUNX2	GDB:392082	CLEIDOCRANIAL DYSPLASIA; CCD CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 1; CBFA1
20 RWS	GDB:9956195	SENSITIVITY
SCA1	GDB:119588	SPINOCEREBELLAR ATAXIA 1; SCA1
SCZD3	GDB:635974	DISORDER-3; SCZD3
SIASD	GDB:433552	SIALIC ACID STORAGE DISEASE; SIASD
25 SOD2	GDB:119597	SUPEROXIDE DISMUTASE 2, MITOCHONDRIAL; SOD2
ST8	GDB:6118456	OVARIAN TUMOR
TAP1	GDB:132668	TRANSPORTER 1, ABC; TAP1
30 TAP2	GDB:132669	TRANSPORTER 2, ABC; TAP2
TFAP2B	GDB:681506	DUCTUS ARTERIOSUS; PDA TRANSCRIPTION FACTOR AP-2 BETA; TFAP2B
TNDM	GDB:9956265	DIABETES MELLITUS, TRANSIENT NEONATAL
35 TNF	GDB:120441	TUMOR NECROSIS FACTOR; TNF

Gene	GDB Accession ID	OMIM Link
5	TPBG	TROPHOBlast GLYCOPROTEIN; TPBG; M6P1
	TPMT	THIOPURINE S-METHYLTRANSFERASE; TPMT
	TULP1	TUBBY-LIKE PROTEIN 1; TULP1
	WISP3	ARTHROPATHY, PROGRESSIVE PSEUDORHEUMATOID, OF CHILDHOOD

10

Table 8: Genes, Locations and Genetic Disorders on Chromosome 7

Gene	GDB Accession ID	OMIM Link
15	AASS	HYPERLYSINEMIA
	ABCB1	P-GLYCOPROTEIN-1; PGY1
	ABCB4	P-GLYCOPROTEIN-3; PGY3
	ACHE	ACETYLCHOLINESTERASE BLOOD GROUP--Yt SYSTEM; YT
20	AQP1	AQUAPORIN-1; AQP1 BLOOD GROUP--COLTON; CO
	ASL	ARGININOSUCCINICACIDURIA
	ASNS	ASPARAGINE SYNTHETASE; ASNS; AS
	AUTS1	DISORDER
25	BPGM	DIPHOSPHOGLYCERATE MUTASE DEFICIENCY OF ERYTHROCYTE
	C7orf2	ACHEIROPODY
	CACNA2D1	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, L TYPE, ALPHA-2/DELTA SUBUNIT; MALIGNANT HYPERTHERMIA SUSCEPTIBILITY-3
	CCM1	CEREBRAL CAVERNOUS MALFORMATIONS 1; CCM1
30	CD36	CD36 ANTIGEN; CD36

Gene	GDB Accession ID	OMIM Link
CFTR	GDB:120584	CYSTIC FIBROSIS; CF DEFERENS, CONGENITAL BILATERAL APLASIA OF; CBAVD; CAVD
5 CHORDOMA	GDB:11498328	
CLCN1	GDB:134688	CHLORIDE CHANNEL 1, SKELETAL MUSCLE; CLCN1
10 CMH6	GDB:9956392	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, WITH WOLFF-PARKINSON-WHITE
CMT2D	GDB:9953232	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, D
15 COL1A2	GDB:119062	COLLAGEN, TYPE I, ALPHA-2 POLYPEPTIDE; COL1A2 OSTEOPENESIS IMPERFECTA TYPE I OSTEOPENESIS IMPERFECTA TYPE IV; OI4
CRS	GDB:119073	CRANIOSYNOSTOSIS, TYPE 1; CRS1
CYMD	GDB:366594	MACULAR EDEMA, CYSTOID
20 DFNAs5	GDB:636174	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 5; DFNAs5
DLD	GDB:120608	LIPOAMIDE DEHYDROGENASE DEFICIENCY, LACTIC ACIDOSIS DUE TO
25 DYT11	GDB:10013754	MYOCLONUS, HEREDITARY ESSENTIAL
EEC1	GDB:136338	ECTRODACTYLY, ECTODERMAL DYSPLASIA, AND CLEFT LIP/PALATE; EEC
30 ELN	GDB:119107	ELASTIN; ELN WILLIAMS-BEUREN SYNDROME; WBS
ETV1	GDB:335229	ETS VARIANT GENE 1; ETV1
FKBP6	GDB:9955215	WILLIAMS-BEUREN SYNDROME; WBS
35 GCK	GDB:127550	DIABETES MELLITUS, AUTOSOMAL DOMINANT, TYPE II GLUCOKINASE; GCK

	Gene	GDB Accession ID	OMIM Link
	GHRHR	GDB:138465	GROWTH HORMONE-RELEASING HORMONE RECEPTOR; GHRHR
5	GHS	GDB:9956363	MICROSOMIA WITH RADIAL DEFECTS
	GLI3	GDB:119990	PALLISTER-HALL SYNDROME; PHS GLI-KRUPPEL FAMILY MEMBER 3; GLI3 POSTAXIAL POLYDACTYLY, TYPE A1 GREIG CEPHALOPOLYSYNDACTYLY SYNDROME; GCPS
10	GPDS1	GDB:9956410	GLAUCOMA, PIGMENT-DISPERSION TYPE
	GUSB	GDB:120025	MUCOPOLYSACCHARIDOSIS TYPE VII
15	HADH	GDB:120033	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
	HLXB9	GDB:136411	HOME BOX GENE HB9; HLXB9 SACRAL AGENESIS, HEREDITARY, WITH PRESACRAL MASS, ANTERIOR MENINGOCELE,
20	HOXA13	GDB:120656	HOME BOX A13; HOXA13
	HPFH2	GDB:128071	HEREDITARY PERSISTENCE OF FETAL HEMOGLOBIN, HETEROCELLULAR, INDIAN
	HRX	GDB:9958999	HRX
25	IAB	GDB:11498909	ANEURYSM, INTRACRANIAL BERRY
	IMMP2L	GDB:11499195	GILLES DE LA TOURETTE SYNDROME; GTS
	KCNH2	GDB:138126	LONG QT SYNDROME, TYPE 2; LQT2
	LAMB1	GDB:119357	LAMININ BETA 1; LAMB1
30	LEP	GDB:136420	LEPTIN; LEP
	MET	GDB:120178	MET PROTO-ONCOGENE; MET
	NCF1	GDB:120222	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-POSITIVE FORM
35	NM	GDB:119454	NEUTROPHIL CHEMOTACTIC RESPONSE; NCR

	Gene	GDB Accession ID	OMIM Link
5	OGDH	GDB:118847	ALPHA-KETOGLUTARATE DEHYDROGENASE DEFICIENCY
	OPN1SW	GDB:119032	TRITANOPIA
	PEX1	GDB:9787110	ZELLWEGER SYNDROME; ZS; PEROXIN-1; PEX1
10	PGAM2	GDB:120280	PHOSPHOGLYCERATE MUTASE, DEFICIENCY OF M SUBUNIT OF
	PMS2	GDB:386406	POSTMEIOTIC SEGREGATION INCREASED (S. CEREVIAE)-2; PMS2
	PON1	GDB:120308	PARAOXONASE 1; PON1
15	PPP1R3A	GDB:136797	PROTEIN PHOSPHATASE 1, REGULATORY (INHIBITOR) SUBUNIT 3; PPP1R3
	PRSS1	GDB:119620	PANCREATITIS, HEREDITARY; PCTT PROTEASE, SERINE, 1; PRSS1
	PTC	GDB:118744	PHENYLTHIOCARBAMIDE TASTING
20	PTPN12	GDB:136846	PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR TYPE, 12; PTPN12
	RP10	GDB:138786	RETINITIS PIGMENTOSA-10; RP10
	RP9	GDB:333931	RETINITIS PIGMENTOSA-9; RP9
25	SERPINE1	GDB:120297	PLASMINOGEN ACTIVATOR INHIBITOR, TYPE I; PAI1
	SGCE	GDB:9958714	MYOCLONUS, HEREDITARY ESSENTIAL
	SHFM1	GDB:128195	SPLIT-HAND/FOOT DEFORMITY, TYPE I; SHFD1
30	SHH	GDB:456309	HOLOPROSENCEPHALY, TYPE 3; HPE3 SONIC HEDGEHOG, DROSOPHILA, HOMOLOG OF; SHH
	SLC26A3	GDB:138165	DOWN-REGULATED IN ADENOMA; DRA CHLORIDE DIARRHEA, FAMILIAL; CLD
	SLC26A4	GDB:5584511	PENDRED SYNDROME; PDS DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 4; DFNB4

Gene	GDB Accession ID	OMIM Link
SLOS	GDB:385950	SMITH-LEMLI-OPITZ SYNDROME
SMAD1	GDB:3763345	SPINAL MUSCULAR ATROPHY, DISTAL, WITH UPPER LIMB PREDOMINANCE; SMAD1
TBXAS1	GDB:128744	THROMBOXANE A SYNTHASE 1; TBXAS1
TWIST	GDB:135694	ACROCEPHALOSYNDACTYLY TYPE III TWIST, DROSOPHILA, HOMOLOG OF; TWIST
ZWS1	GDB:120511	ZELLWEGER SYNDROME; ZS

Table 9: Genes, Locations and Genetic Disorders on Chromosome 8

Gene	GDB Accession ID	OMIM Link
ACHM3	GDB:9120558	PINGELAPESE BLINDNESS
ADRB3	GDB:203869	BETA-3-ADRENERGIC RECEPTOR; ADRB3
ANK1	GDB:118737	SPHEROCYTOSIS, HEREDITARY; HS
CA1	GDB:119047	CARBONIC ANHYDRASE I, ERYTHROCYTE, ELECTROPHORETIC VARIANTS OF; CA1
CA2	GDB:119739	OSTEOPETROSIS WITH RENAL TUBULAR ACIDOSIS
CCAL1	GDB:512892	CHONDROCALCINOSIS WITH EARLY-ONSET OSTEOARTHRITIS; CCAL2
CLN8	GDB:252118	EPILEPSY, PROGRESSIVE, WITH MENTAL RETARDATION; EPMR
CMT4A	GDB:138755	CHARCOT-MARIE-TOOTH NEUROPATHY 4A; CMT4A
CNGB3	GDB:9993286	PINGELAPESE BLINDNESS
COH1	GDB:252122	COHEN SYNDROME; COH1
CPP	GDB:119798	CERULOPLASMIN; CP
CRH	GDB:119804	CORTICOTROPIN-RELEASING HORMONE; CRH
CYP11B1	GDB:120603	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 11-@BETA-HYDROXYLASE DEFICIENCY

	Gene	GDB AccessionID	OMIM Link
5	CYP11B2	GDB:120514	CYTOCHROME P450, SUBFAMILY XIB, POLYPEPTIDE 2; CYP11B2
	DECR1	GDB:453934	2,4-@DIENOYL-CoA REDUCTASE; DECR
	DPYS	GDB:5885803	DIHYDROPYRIMIDINASE; DPYS
	DURS1	GDB:9958126	DUANE SYNDROME
	EBS1	GDB:119856	EPIDERMOLYSIS BULLOSA SIMPLEX, OGNA TYPE
10	ECA1	GDB:10796318	JUVENILE ABSENCE
	EGI	GDB:128830	EPILEPSY, GENERALIZED, IDIOPATHIC; EGI
	EXT1	GDB:135994	EXOSTOSES, MULTIPLE, TYPE I; EXT1 CHONDROSARCOMA
	EYA1	GDB:5215167	BRANCHIOOTORENAL DYSPLASIA EYES ABSENT 1; EYA1
	FGFR1	GDB:119913	ACROCEPHALOSYNDACTYLY TYPE V FIBROBLAST GROWTH FACTOR RECEPTOR-1; FGFR1
15	GNRH1	GDB:133746	GONADOTROPIN-RELEASING HORMONE 1; GNRH1 FAMILIAL HYPOGONADOTROPHIC
	GSR	GDB:119288	GLUTATHIONE REDUCTASE; GSR
	GULOP	GDB:128078	SCURVY
	HR	GDB:595499	ALOPECIA UNIVERSALIS ATRICIA WITH PAPULAR LESIONS HAIRLESS, MOUSE, HOMOLOG OF
	KCNQ3	GDB:9787230	CONVULSIONS, BENIGN FAMILIAL NEONATAL, TYPE 2; BFNC2 POTASSIUM CHANNEL, VOLTAGE-GATED, SUBFAMILY Q, MEMBER 3
20	KFM	GDB:265291	KLIPPEL-FEIL SYNDROME; KFS; KFM
	KWE	GDB:9315120	KERATOLYTIC WINTER ERYTHEMA
	LGCR	GDB:120698	LANGER-GIEDION SYNDROME; LGS
	LPL	GDB:120700	HYPERLIPOPROTEINEMIA, TYPE I
	MCPH1	GDB:9834525	MICROCEPHALY; MCT
35			

Gene	GDB AccessionID	OMIM Link
MOS	GDB:119396	TRANSFORMATION GENE: ONCOGENE MOS
5 MYC	GDB:120208	TRANSFORMATION GENE: ONCOGENE MYC; MYC
NAT1	GDB:125364	ARYLAMIDE ACETYLASE 1; AAC1
NAT2	GDB:125365	ISONIAZID INACTIVATION
10 NBS1	GDB:9598211	NIJMEGEN BREAKAGE SYNDROME
PLAT	GDB:119496	PLASMINOGEN ACTIVATOR, TISSUE; PLAT
PLEC1	GDB:4119073	EPIDERMOLYSIS BULLOSA SIMPLEX AND LIMB-GIRDLE MUSCULAR DYSTROPHY PLECTIN 1; PLEC1
15 PRKDC	GDB:234702	SEVERE COMBINED IMMUNODEFICIENCY DISEASE-1; SCID1 PROTEIN KINASE, DNA-ACTIVATED, CATALYTIC SUBUNIT; PRKDC
PXMP3	GDB:131487	PEROXIN-2; PEX2 ZELLWEGER SYNDROME; ZS
20 RP1	GDB:120352	RETINITIS PIGMENTOSA-1; RP1
SCZD6	GDB:9864736	DISORDER-2; SCZD2
SFTPC	GDB:120373	PULMONARY SURFACTANT APOPROTEIN PSP-C
SGM1	GDB:135350	KLIPPEL-FEIL SYNDROME; KFS; KFM
25 SPG5A	GDB:250332	SPASTIC PARAPLEGIA-5A, AUTOSOMAL RECESSIVE; SPG5A
STAR	GDB:635457	STEROIDOGENIC ACUTE REGULATORY PROTEIN; STAR
TG	GDB:120434	THYROGLOBULIN; TG
30 TRPS1	GDB:594960	TRICORHINOPHALANGEAL SYNDROME, TYPE I; TRPS1
TTPA	GDB:512364	VITAMIN E, FAMILIAL ISOLATED DEFICIENCY OF; VED TOCOPHEROL (ALPHA) TRANSFER PROTEIN; TTPA
35 VMD1	GDB:119631	MACULAR DYSTROPHY, ATYPICAL VITELLIFORM; VMD1

Gene	GDB AccessionID	OMIM Link
WRN	GDB:128446	WERNER SYNDROME; WRN

5 Table 10: Genes, Locations and Genetic Disorders on Chromosome 9

Gene	GDB AccessionID	OMIM Link
ABCA1	GDB:305294	ANALPHALIPOPROTEINEMIA ATP-BINDING CASSETTE 1; ABC1
ABL1	GDB:119640	ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1; ABL1
ABO	GDB:118956	ABO BLOOD GROUP; ABO
ADAMTS13	GDB:9956467	THROMBOCYTOPENIC PURPURA
AK1	GDB:119664	ADENYLYLATE KINASE-1; AK1
15 ALAD	GDB:119665	DELTA-AMINOLEVULINATE DEHYDRATASE; ALAD
ALDH1A1	GDB:119667	ALDEHYDE DEHYDROGENASE-1; ALDH1
ALDOB	GDB:119669	FRUCTOSE INTOLERANCE, HEREDITARY
AMBp	GDB:120696	PROTEIN HC; HCP
20 AMCD1	GDB:437519	ARTHROGRYPOSIS MULTIPLEX CONGENITA, DISTAL, TYPE 1; AMCD1
ASS	GDB:119010	CITRULLINEMIA
BDMF	GDB:9954424	BONE DYSPLASIA WITH MEDULLARY FIBROSARCOMA
25 BSCL	GDB:9957720	SEIP SYNDROME
C5	GDB:119734	COMPLEMENT COMPONENT-5, DEFICIENCY OF
CDKN2A	GDB:335362	MELANOMA, CUTANEOUS MALIGNANT, 2; CMM2 CYCLIN-DEPENDENT KINASE INHIBITOR 2A; CDKN2A
30 CHAC	GDB:6268491	CHOREOACANTHOCYTOSIS; CHAC
CHH	GDB:138268	CARTILAGE-HAIR HYPOPLASIA; CHH
CMD1B	GDB:677147	CARDIOMYOPATHY, DILATED 1B; CMD1B
35 COL5A1	GDB:131457	COLLAGEN, TYPE V, ALPHA-1 POLYPEPTIDE; COL5A1

	Gene	GDB AccessionID	OMIM Link
	CRAT	GDB:359759	CARNITINE ACETYLTRANSFERASE; CRAT
5	DBH	GDB:119836	DOPAMINE BETA-HYDROXYLASE, PLASMA; DBH
	DFNB11	GDB:1220180	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 7; DFNB7
	DFNB7	GDB:636178	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 7; DFNB7
10	DNAI1	GDB:11500297	IMMOTILE CILIA SYNDROME-1; ICS1
	DYS	GDB:137085	DYSAUTONOMIA, FAMILIAL; DYS
	DYT1	GDB:119854	DYSTONIA 1, TORSION; DYT1
	ENG	GDB:137193	ENDOGLIN; ENG
15	EPB72	GDB:128993	ERYTHROCYTE SURFACE PROTEIN BAND 7.2; EPB72 STOMATOCYTOSIS I
	FANCC	GDB:132672	FANCONI ANEMIA, COMPLEMENTATION GROUP C; FACC
	FBP1	GDB:141539	FRUCTOSE-1,6-BISPHOPHATASE 1; FBP1
20	FCMD	GDB:250412	FUKUYAMA-TYPE CONGENITAL MUSCULAR DYSTROPHY; FCMD
	FRDA	GDB:119951	FRIEDREICH ATAXIA 1; FRDA1
	GALT	GDB:119971	GALACTOSEMIA
	GLDC	GDB:128611	HYPERGLYCINEMIA, ISOLATED NONKETOTIC, TYPE I; NKH1
25	GNE	GDB:9954891	INCLUSION BODY MYOPATHY; IBM2
	GSM1	GDB:9784210	GENIOSPASM 1; GSM1
	GSN	GDB:120019	AMYLOIDOSIS V GELSOLIN; GSN
30	HSD17B3	GDB:347487	PSEUDOHERMAPHRODITISM, MALE, WITH GYNECOMASTIA
	HSN1	GDB:3853677	NEUROPATHY, HEREDITARY SENSORY, TYPE 1
	IBM2	GDB:3801447	INCLUSION BODY MYOPATHY; IBM2
35	LALL	GDB:9954426	LEUKEMIA, ACUTE, WITH LYMPHOMATOUS FEATURES; LALL

Gene	GDB AccessionID	OMIM Link
LCCS	GDB:386141	LETHAL CONGENITAL CONTRACTURE SYNDROME; LCCS
5	LGMD2H	DYSTROPHY, HUTTERITE TYPE
	LMX1B	NAIL-PATELLA SYNDROME; NPS1
	MLLT3	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 3; MLLT3
	MROS	MELKERSSON SYNDROME
10	MSSE	EPITHELIOMA, SELF-HEALING SQUAMOUS
	NOTCH1	NOTCH, DROSOPHILA, HOMOLOG OF, 1; NOTCH1
	ORM1	OROSOMUCOID 1; ORM1
15	PAPPA	PREGNANCY-ASSOCIATED PLASMA PROTEIN A; PAPPA
	PIP5K1B	FRIEDREICH ATAXIA 1; FRDA1
	PTCH	BASAL CELL NEVUS SYNDROME; BCNS PATCHED, DROSOPHILA, HOMOLOG OF; PTCH
20	PTGS1	PROSTAGLANDIN-ENDOPEROXIDASE SYNTHASE 1; PTGS1
	RLN1	RELAXIN; RLN1
	RLN2	RELAXIN, OVARIAN, OF PREGNANCY
25	RMRP	MITOCHONDRIAL RNA-PROCESSING ENDORIBONUCLEASE, RNA COMPONENT OF; RMRP; CARTILAGE-HAIR HYPOPLASIA; CHH
	ROR2	BRACHYDACTYLY, TYPE B; BDB ROBINOW SYNDROME, RECESSIVE FORM NEUROTROPHIC TYROSINE KINASE, RECEPTOR-RELATED 2; NTRK2
30	RPD1	RETINITIS PIGMENTOSA-DEAFNESS SYNDROME 1, AUTOSOMAL DOMINANT
	SARDH	SARCOSINEMIA
	TDFA	FACTOR, AUTOSOMAL

Gene	GDB AccessionID	OMIM Link
TEK	GDB:344185	VENOUS MALFORMATIONS, MULTIPLE CUTANEOUS AND MUCOSAL; VMCM TEK TYROSINE KINASE, ENDOTHELIAL; TEK
5	TSC1	TUBEROUS SCLEROSIS-1; TSC1
	TYRP1	TYROSINASE-RELATED PROTEIN 1; TYRP1 ALBINISM III XANTHISM
	XPA	XERODERMA PIGMENTOSUM I

10 Table 11: Genes, Locations and Genetic Disorders on Chromosome 10

Gene	GDB Accession ID	OMIM Link
CACNB2	GDB:132014	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, BETA-2 SUBUNIT; CACNB2
15	COL17A1	COLLAGEN, TYPE XVII, ALPHA-1 POLYPEPTIDE; COL17A1
	CUBN	MEGALOBLASTIC ANEMIA 1; MGA1
	CYP17	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 17-ALPHA-HYDROXYLASE DEFICIENCY
	CYP2C19	CYTOCHROME P450, SUBFAMILY IIC, POLYPEPTIDE 19; CYP2C19
20	CYP2C9	CYTOCHROME P450, SUBFAMILY IIC, POLYPEPTIDE 9; CYP2C9
	EGR2	EARLY GROWTH RESPONSE-2; EGR2
	EMX2	EMPTY SPIRACLES, DROSOPHILA, 2, HOMOLOG OF; EMX2
	EPT	EPILEPSY, PARTIAL; EPT
25	ERCC6	EXCISION-REPAIR CROSS-COMPLEMENTING RODENT REPAIR DEFICIENCY, COMPLEMENTATION
	FGFR2	ACROCEPHALOSYNDACTYLY TYPE V FIBROBLAST GROWTH FACTOR RECEPTOR-2; FGFR2
30		
35		

	Gene	GDB Accession ID	OMIM Link
5	HK1	GDB:120044	HEXOKINASE-1; HK1
	HOX11	GDB:119607	HOME BOX-11; HOX11
	HPS	GDB:127359	HERMANSKY-PUDLAK SYNDROME; HPS
	IL2RA	GDB:119345	INTERLEUKIN-2 RECEPTOR, ALPHA; IL2RA
10	LGII	GDB:9864936	EPILEPSY, PARTIAL; EPT
	LIPA	GDB:120153	WOLMAN DISEASE
	MAT1A	GDB:129077	METHIONINE ADENOSYLTRANSFERASE DEFICIENCY
	MBL2	GDB:120167	MANNOSE-BINDING PROTEIN, SERUM; MBP1
15	MKI67	GDB:120185	PROLIFERATION-RELATED Ki-67 ANTIGEN; MKI67
	MXII1	GDB:137182	MAX INTERACTING PROTEIN 1; MXII1
	OAT	GDB:120246	ORNITHINE AMINOTRANSFERASE DEFICIENCY
	OATL3	GDB:215803	ORNITHINE AMINOTRANSFERASE DEFICIENCY
20	PAX2	GDB:138771	PAIRED BOX HOMEOTIC GENE 2; PAX2
	PCBD	GDB:138478	PTERIN-4-ALPHA-CARBINOLAMINE DEHYDRATASE; PCBD PRIMAPTERINURIA
	PEO1	GDB:632784	PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA; PEO
	PHYH	GDB:9263423	REFSUM DISEASE PHYTANOYL-CoA HYDROXYLASE; PHYH
25	PNLIP	GDB:127916	LIPASE, CONGENITAL ABSENCE OF PANCREATIC
	PSAP	GDB:120366	PROSAPOSIN; PSAP
35			

Gene	GDB Accession ID	OMIM Link
PTEN	GDB:6022948	MACROCEPHALY, MULTIPLE LIPOSAS AND HEMANGIOMATA MULTIPLE HAMARTOMA SYNDROME; MHAM POLYPOSIS, JUVENILE INTESTINAL PHOSPHATASE AND TENSIN HOMOLOG; PTEN
RBP4	GDB:120342	RETINOL-BINDING PROTEIN, PLASMA; RBP4
RDPA	GDB:9954445	REFSUM DISEASE WITH INCREASED PIPECOLICACIDEMIA; RDPA
RET	GDB:120346	RET PROTO-ONCOGENE; RET
SDF1	GDB:433267	STROMAL CELL-DERIVED FACTOR 1; SDF1
SFTPA1	GDB:119593	PULMONARY SURFACTANT APOPROTEIN PSP-A; PSAP
SFTPD	GDB:132674	PULMONARY SURFACTANT APOPROTEIN PSP-D; PSP-D
SHFM3	GDB:386030	SPLIT-HAND/FOOT MALFORMATION, TYPE 3; SHFM3
SIAL	GDB:6549924	NEURAMINIDASE DEFICIENCY
THC2	GDB:10794765	THROMBOCYTOPENIA
TNFRSF6	GDB:132671	APOPTOSIS ANTIGEN 1; APT1
UFS	GDB:6380714	UROFACIAL SYNDROME; UFS
UROS	GDB:128112	PORPHYRIA, CONGENITAL ERYTHROPOIETIC; CEP

Table 12: Genes, Locations and Genetic Disorders on Chromosome 11

Gene	GDB Accession ID	OMIM Link
AA	GDB:568984	ATROPHIA AREATA; AA
ABCC8	GDB:591370	SULFONYLUREA RECEPTOR; SUR PERSISTENT HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY
ACAT1	GDB:126861	ALPHA-METHYLACETOACETICACIDURIA
ALX4	GDB:10450304	PARIELAL FORAMINA, SYMMETRIC; PFM

Gene	GDB Accession ID	OMIM Link
AMPD3	GDB:136013	ADENOSINE MONOPHOSPHATE DEAMINASE-3; AMPD3
5 ANC	GDB:9954484	CANAL CARCINOMA
APOA1	GDB:119684	AMYLOIDOSIS, FAMILIAL VISCERAL APOLIPOPROTEIN A-I OF HIGH DENSITY LIPOPROTEIN; APOA1
10 APOA4	GDB:119000	APOLIPOPROTEIN A-IV; APOA4
APOC3	GDB:119001	APOLIPOPROTEIN C-III; APOC3
ATM	GDB:593364	ATAXIA-TELANGIECTASIA; AT
BSCL2	GDB:9963996	SEIP SYNDROME
BWS	GDB:120567	BECKWITH-WIEDEMANN SYNDROME; BWS
15 CALCA	GDB:120571	CALCITONIN/CALCITONIN-RELATED POLYPEPTIDE, ALPHA; CALCA
CAT	GDB:119049	CATALASE; CAT
CCND1	GDB:128222	LEUKEMIA, CHRONIC LYMPHATIC; CLL CYCLIN D1; CCND1
20 CD3E	GDB:119764	CD3E ANTIGEN, EPSILON POLYPEPTIDE; CD3E
CD3G	GDB:119765	T3 T-CELL ANTIGEN, GAMMA CHAIN; T3G; CD3G
25 CD59	GDB:119769	CD59 ANTIGEN P18-20; CD59 HUMAN LEUKOCYTE ANTIGEN MIC11; MIC11
CDKN1C	GDB:593296	CYCLIN-DEPENDENT KINASE INHIBITOR 1C; CDKN1C
CLN2	GDB:125228	CEROID-LIPOFUSCINOSIS, NEURONAL 2, LATE INFANTILE TYPE; CLN2
30 CNTF	GDB:125919	CILIARY NEUROTROPHIC FACTOR; CNTF
CPT1A	GDB:597642	HYPOGLYCEMIA, HYPOKETOTIC, WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE CARNITINE PALMITOYLTRANSFERASE I, LIVER; CPT1A

Gene	GDB Accession ID	OMIM Link
CTSC	GDB:642234	KERATOSIS PALMOPLANTARIS WITH PERIODONTOPATHIA KERATOSIS PALMOPLANTARIS WITH PERIODONTOPATHIA AND ONYCHOGRYPHOsis CATHEPSIN C; CTSC
DDB1	GDB:595014	DNA DAMAGE-BINDING PROTEIN; DDB1
DDB2	GDB:595015	DNA DAMAGE-BINDING PROTEIN-2; DDB2
DHCR7	GDB:9835302	SMITH-LEMLI-OPITZ SYNDROME
DLAT	GDB:118785	CIRRHOSIS, PRIMARY; PBC
DRD4	GDB:127782	DOPAMINE RECEPTOR D4; DRD4
ECB2	GDB:9958955	POLYCYTHEMIA, BENIGN FAMILIAL
ED4	GDB:9837373	DYSPLASIA, MARGARITA TYPE
EVR1	GDB:134029	EXUDATIVE VITREORETINOPATHY, FAMILIAL; EVR EXT2GDB:344921EXOSTOSES, MULTIPLE, TYPE II; EXT2 CHONDROSARCOMA
F2	GDB:119894	COAGULATION FACTOR II; F2
FSHB	GDB:119955	FOLLICLE-STIMULATING HORMONE, BETA POLYPEPTIDE; FSHB
FTH1	GDB:120617	FERRITIN HEAVY CHAIN 1; FTH1
GIF	GDB:118800	PERNICIOUS ANEMIA, CONGENITAL, DUE TO DEFECT OF INTRINSIC FACTOR
GSD1B	GDB:9837619	GLYCOGEN STORAGE DISEASE Ib
GSD1C	GDB:9837637	STORAGE DISEASE Ic
HBB	GDB:119297	HEMOGLOBIN--BETA LOCUS; HBB
HBBP1	GDB:120035	HEMOGLOBIN--BETA LOCUS; HBB
HBD	GDB:119298	HEMOGLOBIN--DELTA LOCUS; HBD
HBE1	GDB:119299	HEMOGLOBIN--EPSILON LOCUS; HBE1
HBG1	GDB:119300	HEMOGLOBIN, GAMMA A; HBG1
HBG2	GDB:119301	HEMOGLOBIN, GAMMA G; HBG2
HMBS	GDB:120528	PORPHYRIA, ACUTE INTERMITTENT; AIP
HND	GDB:9954478	HARTNUP DISORDER

Gene	GDB Accession ID	OMIM Link
HOMG2	GDB:9956484	MAGNESIUM WASTING, RENAL
HRAS	GDB:120684	BLADDER CANCER V-HA-RAS HARVEY RAT SARCOMA VIRAL ONCOGENE HOMOLOG; HRAS
HVBS1	GDB:120069	CANCER, HEPATOCELLULAR
10 IDDM2	GDB:128530	DIABETES MELLITUS, INSULIN-DEPENDENT, 2 DIABETES MELLITUS, JUVENILE-ONSET INSULIN-DEPENDENT; IDDM
IGER	GDB:119696	IgE RESPONSIVENESS, ATOPIC; IGER
INS	GDB:119349	INSULIN; INS
JBS	GDB:120111	JACOBSEN SYNDROME; JBS
15 KCNJ11	GDB:7009893	POTASSIUM CHANNEL, INWARDLY-RECTIFYING, SUBFAMILY J, MEMBER 11; KCNJ11 PERSISTENT HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY
20 KCNJ1	GDB:204206	POTASSIUM CHANNEL, INWARDLY-RECTIFYING, SUBFAMILY J, MEMBER 1; KCNJ1
KCNQ1	GDB:741244	LONG QT SYNDROME, TYPE 1; LQT1
LDHA	GDB:120141	LACTATE DEHYDROGENASE-A; LDHA
25 LRP5	GDB:9836818	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME; OPPG HIGH BONE MASS
MEN1	GDB:120173	MULTIPLE ENDOCRINE NEOPLASIA, TYPE 1; MEN1
30 MLL	GDB:128819	MYELOID/LYMPHOID OR MIXED-LINEAGE LEUKEMIA; MLL
MTACR1	GDB:125743	MULTIPLE TUMOR ASSOCIATED CHROMOSOME REGION 1; MTACR1
MYBPC3	GDB:579615	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 4; CMH4 MYOSIN-BINDING PROTEIN C, CARDIAC; MYBPC3

Gene	GDB Accession ID	OMIM Link
5 MYO7A	GDB:132543	MYOSIN VIIA; MYO7A DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 2; DFNB2 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 11; DFNA11
10 NNO1	GDB:10450513	SIMPLE, AUTOSOMAL DOMINANT
15 OPPG	GDB:3789438	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME; OPPG
20 OPTB1	GDB:9954474	OSTEOPETROSIS, AUTOSOMAL RECESSIVE
PAX6	GDB:118997	PAIRED BOX HOMEOTIC GENE 6; PAX6
PC	GDB:119472	PYRUVATE CARBOXYLASE DEFICIENCY
15 PDX1	GDB:9836634	PYRUVATE DEHYDROGENASE COMPLEX, COMPONENT X
PGL2	GDB:511177	PARAGANGLIOMAS, FAMILIAL NONCHROMAFFIN, 2; PGL2
PGR	GDB:119493	PROGESTERONE RESISTANCE
20 PORC	GDB:128610	PORPHYRIA, CHESTER TYPE; PORC
PTH	GDB:119522	PARATHYROID HORMONE; PTH
PTS	GDB:118856	6-@PYRUVYLtetrahydropterin SYNTHASE; PTS
25 PVRL1	GDB:583951	ECTODERMAL DYSPLASIA, CLEFT LIP AND PALATE, HAND AND FOOT DEFORMITY, DYSPLASIA, MARGARITA TYPE POLIOVIRUS RECEPTOR RELATED; PVRR
PYGM	GDB:120329	GLYCOGEN STORAGE DISEASE V
30 RAG1	GDB:120334	RECOMBINATION ACTIVATING GENE-1; RAG1
RAG2	GDB:125186	RECOMBINATION ACTIVATING GENE-2; RAG2
ROM1	GDB:120350	ROD OUTER SEGMENT PROTEIN-1; ROM1
35 SAA1	GDB:120364	SERUM AMYLOID A1; SAA1
SCA5	GDB:378219	SPINOCEREBELLAR ATAXIA 5; SCA5

Gene	GDB Accession ID	OMIM Link
SCZD2	GDB:118874	DISORDER-2; SCZD2
SDHD	GDB:132456	PARAGANGLIOMAS, FAMILIAL NONCHROMAFFIN, 1; PGL1
SERPING1	GDB:119041	ANGIONEUROTIC EDEMA, HEREDITARY; HANE
SMPD1	GDB:128144	NIEMANN-PICK DISEASE
TCIRG1	GDB:9956269	OSTEOPETROSIS, AUTOSOMAL RECESSIVE
TCL2	GDB:9954468	LEUKEMIA, ACUTE T-CELL; ATL
TECTA	GDB:6837718	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 8; DFNA8 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 12; DFNA12
TH	GDB:119612	TYROSINE HYDROXYLASE; TH
TREH	GDB:9958953	TREHALASE
TSG101	GDB:1313414	TUMOR SUSCEPTIBILITY GENE 101; TSG101
TYR	GDB:120476	ALBINISM I
USH1C	GDB:132544	USHER SYNDROME, TYPE IC; USH1C
VMD2	GDB:133795	VITELLIFORM MACULAR DYSTROPHY; VMD2
VRNI	GDB:135662	VITREORETINOPATHY, NEOVASCULAR INFLAMMATORY; VRNI
WT1	GDB:120496	FRASIER SYNDROME WILMS TUMOR; WT1
WT2	GDB:118886	MULTIPLE TUMOR ASSOCIATED CHROMOSOME REGION 1; MTACR1
ZNF145	GDB:230064	PROMYELOCYTIC LEUKEMIA ZINC FINGER; PLZF

Table 13: Genes, Locations and Genetic Disorders on Chromosome 12

Gene	GDB Accession ID	OMIM Link
A2M	GDB:119639	ALPHA-2-MACROGLOBULIN; A2M

	Gene	GDB Accession ID	OMIM Link
	AAAS	GDB:9954498	GLUCOCORTICOID DEFICIENCY AND ACHALASIA
5	ACADS	GDB:118959	ACYL-CoA DEHYDROGENASE, SHORT-CHAIN; ACADS
	ACLS	GDB:136346	ACROCALLOSAL SYNDROME; ACLS
10	ACVRL1	GDB:230240	OSLER-RENDU-WEBER SYNDROME 2; ORW2 ACTIVIN A RECEPTOR, TYPE II-LIKE KINASE 1; ACVRL1
	ADHR	GDB:9954488	VITAMIN D-RESISTANT RICKETS, AUTOSOMAL DOMINANT
	ALDH2	GDB:119668	ALDEHYDE DEHYDROGENASE-2; ALDH2
15	AMHR2	GDB:696210	ANTI-MULLERIAN HORMONE TYPE II RECEPTOR; AMHR2
	AOM	GDB:118998	STICKLER SYNDROME, TYPE I; STL1
20	AQP2	GDB:141853	AQUAPORIN-2; AQP2 DIABETES INSIPIDUS, RENAL TYPE DIABETES INSIPIDUS, RENAL TYPE, AUTOSOMAL RECESSIVE
	ATD	GDB:696353	ASPHYXIATING THORACIC DYSTROPHY; ATD
	ATP2A2	GDB:119717	ATPase, Ca(2+)-TRANSPORTING, SLOW-TWITCH; ATP2A2 DARIER-WHITE DISEASE; DAR
25	BDC	GDB:5584359	BRACHYDACTYLY, TYPE C; BDC
	C1R	GDB:119729	COMPLEMENT COMPONENT-C1r, DEFICIENCY OF
	CD4	GDB:119767	T-CELL ANTIGEN T4/LEU3; CD4
30	CDK4	GDB:204022	CYCLIN-DEPENDENT KINASE 4; CDK4
	CNA1	GDB:252119	CORNEA PLANA 1; CNA1
35	COL2A1	GDB:119063	STICKLER SYNDROME, TYPE I; STL1 COLLAGEN, TYPE II, ALPHA-1 CHAIN; COL2A1 ACHONDROGENESIS, TYPE II; ACG2

	Gene	GDB Accession ID	OMIM Link
	CYP27B1	GDB:9835730	PSEUDOVITAMIN D DEFICIENCY RICKETS; PDDR
5	DRPLA	GDB:270336	DENTATORUBRAL-PALLIDOLUYSIAN ATROPHY; DRPLA
	ENUR2	GDB:666422	ENURESIS, NOCTURNAL, 2; ENUR2
	FEOM1	GDB:345037	FIBROSIS OF EXTRAOCULAR MUSCLES, CONGENITAL; FEOM
10	FPF	GDB:9848880	PERIODIC FEVER, AUTOSOMAL DOMINANT
	GNB3	GDB:120005	GUANINE NUCLEOTIDE-BINDING PROTEIN, BETA POLYPEPTIDE-3; GNB3
15	GNS	GDB:120006	MUCOPOLYSACCHARIDOSIS TYPE IIID
	HAL	GDB:120746	HISTIDINEMIA
	HBPI	GDB:701889	BRACHYDACTYLY WITH HYPERTENSION
20	HMGIC	GDB:362658	HIGH MOBILITY GROUP PROTEIN ISOFORM I-C; HMGIC
	HMN2	GDB:9954508	MUSCULAR ATROPHY, ADULT SPINAL
	HPD	GDB:135978	TYROSINEMIA, TYPE III
25	IGF1	GDB:120081	INSULINLIKE GROWTH FACTOR 1; IGF1
	KCNA1	GDB:127903	POTASSIUM VOLTAGE-GATED CHANNEL, SHAKER-RELATED SUBFAMILY, MEMBER
	KERA	GDB:252121	CORNEA PLANA 2; CNA2
30	KRAS2	GDB:120120	V-KI-RAS2 KIRSTEN RAT SARCOMA 2 VIRAL ONCOGENE HOMOLOG; KRAS2
	KRT1	GDB:128198	KERATIN 1; KRT1
	KRT2A	GDB:407640	ICHTHYOSIS, BULLOUS TYPE KERATIN 2A; KRT2A
35	KRT3	GDB:136276	KERATIN 3; KRT3

	Gene	GDB Accession ID	OMIM Link
	KRT4	GDB:120697	KERATIN 4; KRT4
5	KRT5	GDB:128110	EPIDERMOLYSIS BULLOSA HERPETIFORMIS, DOWLING-MEARA TYPE KERATIN 5; KRT5
	KRT6A	GDB:128111	KERATIN 6A; KRT6A
10	KRT6B	GDB:128113	KERATIN 6B; KRT6B PACHYONYCHIA CONGENITA, JACKSON-LAWLER TYPE
	KRTHB6	GDB:702078	MONILETHRIX KERATIN, HAIR BASIC (TYPE II) 6
15	LDHB	GDB:120147	LACTATE DEHYDROGENASE-B; LDHB
	LYZ	GDB:120160	AMYLOIDOSIS, FAMILIAL VISCERAL LYSOZYME; LYZ
	MGCT	GDB:9954504	TESTICULAR TUMORS
	MPE	GDB:120191	MALIGNANT PROLIFERATION OF
	MVK	GDB:134189	MEVALONICACIDURIA
20	MYL2	GDB:128829	MYOSIN, LIGHT CHAIN, REGULATORY VENTRICULAR; MYL2
	NS1	GDB:439388	NOONAN SYNDROME 1; NS1
	OAP	GDB:120245	OSTEOARTHROSIS, PRECOCIOUS; OAP
25	PAH	GDB:119470	PHENYLKETONURIA; PKU1
	PPKB	GDB:696352	PALMOPLANTAR KERATODERMA, BOTHNIAN TYPE; PPKB
	PRB3	GDB:119513	PAROTID SALIVARY GLYCOPROTEIN; G1
30	PXR1	GDB:433739	ZELLWEGER SYNDROME; ZS PEROXISOME RECEPTOR 1; PXR1
	RLS	GDB:11501392	ACROMELALGIA, HEREDITARY
	RSN	GDB:139158	RESTIN; RSN
35	SAS	GDB:128054	SARCOMA AMPLIFIED SEQUENCE; SAS

	<b>Gene</b>	<b>GDB Accession ID</b>	<b>OMIM Link</b>
5	SCA2	GDB:128034	SPINOCEREBELLAR ATAXIA 2; SCA2 ATAXIN-2; ATX2
	SCNN1A	GDB:366596	SODIUM CHANNEL, NONVOLTAGE-GATED, 1; SCNN1A
	SMAL	GDB:9954506	SPINAL MUSCULAR ATROPHY, CONGENITAL NONPROGRESSIVE, OF LOWER LIMBS
10	SPPM	GDB:9954502	SCAPULOPERONEAL MYOPATHY; SPM
	SPSMA	GDB:9954510	SCAPULOPERONEAL AMYOTROPHY, NEUROGENIC, NEW ENGLAND TYPE
15	TBX3	GDB:681969	ULNAR-MAMMARY SYNDROME; UMS T-BOX 3; TBX3
	TBX5	GDB:6175917	HOLT-ORAM SYNDROME; HOS T-BOX 5; TBX5
20	TCF1	GDB:125297	TRANSCRIPTION FACTOR 1, HEPATIC; TCF1 MATURITY-ONSET DIABETES OF THE YOUNG, TYPE III; MODY3
	TPI1	GDB:119617	TRIOSEPHOSPHATE ISOMERASE 1; TPI1
25	TSC3	GDB:127930	SCLEROSIS-3; TSC3
	ULR	GDB:594089	UTERINE
30	VDR	GDB:120487	VITAMIN D-RESISTANT RICKETS WITH END-ORGAN UNRESPONSIVENESS TO 1,25-DIHYDROXYCHOLECALCIFEROL VITAMIN D RECEPTOR; VDR
	VWF	GDB:119125	VON WILLEBRAND DISEASE; VWD

30 Table 14: Genes, Locations and Genetic Disorders on Chromosome 13

<b>Gene</b>	<b>GDB Accession ID</b>	<b>OMIM Link</b>
35	ATP7B	WILSON DISEASE; WND
	BRCA2	BREAST CANCER 2, EARLY-ONSET; BRCA2

Gene	GDB Accession ID	OMIM Link
5 BRCD1	GDB:9954522	BREAST CANCER, DUCTAL, 1; BRCD1
CLNS	GDB:230991	CEROID-LIPOFUSCOSIS, NEURONAL 5; CLNS
CPB2	GDB:129546	CARBOXYPEPTIDASE B2, PLASMA; CPB2
ED2	GDB:9834522	ECTODERMAL DYSPLASIA, HIDROTIC; HED
10 EDNRB	GDB:129075	ENDOTHELIN-B RECEPTOR; EDNRB HIRSCHSPRUNG DISEASE-2; HSCR2
ENUR1	GDB:594516	ENURESIS, NOCTURNAL, 1; ENUR1
ERCC5	GDB:120515	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 5; ERCC5
15 F10	GDB:119890	X, QUANTITATIVE VARIATION IN FACTOR X DEFICIENCY; F10
F7	GDB:119897	FACTOR VII DEFICIENCY
20 GJB2	GDB:125247	GAP JUNCTION PROTEIN, BETA-2, 26 KD; GJB2 DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 1; DFNB1 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 3; DFNA3
25 GJB6	GDB:9958357	ECTODERMAL DYSPLASIA, HIDROTIC; HED DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 3; DFNA3
IPF1	GDB:448899	INSULIN PROMOTER FACTOR 1; IPF1
MBS1	GDB:128365	MOEBIUS SYNDROME; MBS
MCOR	GDB:9954520	CONGENITAL
30 PCCA	GDB:119473	GLYCINEMIA, KETOTIC, I
RB1	GDB:118734	BLADDER CANCER RETINOBLASTOMA; RB1
RHOK	GDB:371598	RHODOPSIN KINASE; RHOK
SCZD7	GDB:9864734	DISORDER-2; SCZD2
35 SGCG	GDB:3763329	MUSCULAR DYSTROPHY, LIMB GIRDLE, TYPE 2C; LGMD2C

Gene	GDB Accession ID	OMIM Link
SLC10A2	GDB:677534	SOLUTE CARRIER FAMILY 10, MEMBER 2; SLC10A2
SLC25A15	GDB:120042	HYPERNORNITHINEMIA-HYPERAMMONEMIA-HOMOCITRULLINURIA SYNDROME
5 STARP1	GDB:635459	STEROIDOGENIC ACUTE REGULATORY PROTEIN; STAR
ZNF198	GDB:6382650	ZINC FINGER PROTEIN-198; ZNF198

Table 15: Genes, Locations and Genetic Disorders on Chromosome 14

Gene	GDB Accession ID	OMIM Link
ACHM1	GDB:132458	COLORBLINDNESS, TOTAL
ARVD1	GDB:371339	ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 1; ARVD1
15 CTAA1	GDB:265299	CATARACT, ANTERIOR POLAR 1; CTAA1
DAD1	GDB:407505	DEFENDER AGAINST CELL DEATH; DAD1
DFNB5	GDB:636176	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 5; DFNB5
20 EML1	GDB:6328385	USHER SYNDROME, TYPE IA; USH1A
GALC	GDB:119970	KRABBE DISEASE
GCH1	GDB:118798	DYSTONIA, PROGRESSIVE, WITH DIURNAL VARIATION GTP CYCLOHYDROLASE I DEFICIENCY GTP CYCLOHYDROLASE I; GCH1
25 HE1	GDB:9957680	MALFORMATIONS, MULTIPLE, WITH LIMB ABNORMALITIES AND HYPOPITUITARISM
IBGC1	GDB:10450404	CEREBRAL CALCIFICATION, NONARTERIOSCLEROTIC

30

35

Gene	GDB Accession ID	OMIM Link
IGH@	GDB:118731	IgA CONSTANT HEAVY CHAIN 1; IGHA1 IMMUNOGLOBULIN: D (DIVERSITY) REGION OF HEAVY CHAIN IgA CONSTANT HEAVY CHAIN 2; IGHA2 IMMUNOGLOBULIN: J (JOINING) LOCI OF HEAVY CHAIN; IGHJ IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1 IMMUNOGLOBULIN: VARIABLE REGION OF HEAVY CHAINS--Hv1; IGHV IgG HEAVY CHAIN LOCUS; IGHG1 IMMUNOGLOBULIN Gm-3; IGHG2 IMMUNOGLOBULIN Gm-2; IGHG3 IMMUNOGLOBULIN Gm-4; IGHG4 IMMUNOGLOBULIN: HEAVY DELTA CHAIN; IGHD IMMUNOGLOBULIN: HEAVY EPSILON CHAIN; IGHE
IGHC group	GDB:9992632	IgA CONSTANT HEAVY CHAIN 1; IGHA1 IgA CONSTANT HEAVY CHAIN 2; IGHA2 IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1 IgG HEAVY CHAIN LOCUS; IGHG1 IMMUNOGLOBULIN Gm-2; IGHG2 IMMUNOGLOBULIN Gm-3; IGHG3 IMMUNOGLOBULIN Gm-4; IGHG4 IMMUNOGLOBULIN: HEAVY DELTA CHAIN; IGHD IMMUNOGLOBULIN: HEAVY EPSILON CHAIN; IGHE
IGHG1	GDB:120085	IgG HEAVY CHAIN LOCUS; IGHG1
IGHM	GDB:120086	IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1
IGHR	GDB:9954529	G1(A1) SYNDROME
IV	GDB:139274	INVERSUS VISCIERUM
LTBP2	GDB:453890	LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 2; LTBP2
MCOP	GDB:9954527	MICROPHTHALMOS
MJD	GDB:118840	MACHADO-JOSEPH DISEASE; MJD
MNG1	GDB:6540062	GOITER, MULTINODULAR 1; MNG1
MPD1	GDB:230271	MYOPATHY, LATE DISTAL HEREDITARY
MPS3C	GDB:9954532	MUCOPOLYSACCHARIDOSIS TYPE IIIC
MYH6	GDB:120214	MYOSIN, HEAVY POLYPEPTIDE 6; MYH6

	Gene	GDB Accession ID	OMIM Link
5	MYH7	GDB:120215	MYOSIN, CARDIAC, HEAVY CHAIN, BETA; MYH7
	NP	GDB:120239	NUCLEOSIDE PHOSPHORYLASE; NP
	PABPN1	GDB:567135	OCULOPHARYNGEAL MUSCULAR DYSTROPHY; OPMD OCULOPHARYNGEAL MUSCULAR DYSTROPHY, AUTOSOMAL RECESSIVE POLYADENYLATE-BINDING PROTEIN-2; PABP2
10	PSEN1	GDB:135682	ALZHEIMER DISEASE, FAMILIAL, TYPE 3; AD3
	PYGL	GDB:120328	GLYCOGEN STORAGE DISEASE VI
15	RPGRIP1	GDB:11498766	AMAUROSIS CONGENITA OF LEBER I
	SERPINA1	GDB:120289	PROTEASE INHIBITOR 1; PI
	SERPINA3	GDB:118955	ALPHA-1-ANTICHYMOTRYPSIN; AACT
	SERPINA6	GDB:127865	CORTICOSTEROID-BINDING GLOBULIN; CBG
20	SLC7A7	GDB:9863033	DIBASICAMINOACIDURIA II
	SPG3A	GDB:230126	SPASTIC PARAPLEGIA-3, AUTOSOMAL DOMINANT; SPG3A
	SPTB	GDB:119602	ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE HEREDITARY HEMOLYTIC SPECTRIN, BETA, ERYTHROCYTIC; SPTB
25	TCL1A	GDB:250785	T-CELL LYMPHOMA OR LEUKEMIA
	TCRAV17S1	GDB:642130	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA
	TCRAV5S1	GDB:451966	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA
30	TGM1	GDB:125299	TRANSGlutaminase 1; TGM1 ICHTHYOSIS CONGENITA
	TITF1	GDB:132588	THYROID TRANSCRIPTION FACTOR 1; TITF1
	TMIP	GDB:9954523	AND ULNA, DUPLICATION OF, WITH ABSENCE OF TIBIA AND RADIUS
35	TRA@	GDB:120404	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA

Gene	GDB Accession ID	OMIM Link
TSHR	GDB:125313	THYROTROPIN, UNRESPONSIVENESS TO
USH1A	GDB:118885	USHER SYNDROME, TYPE IA; USH1A
VP	GDB:120492	PORPHYRIA VARIEGATA

5 Table 16: Genes, Locations and Genetic Disorders on Chromosome 15

Gene	GDB Accession ID	OMIM Link
ACCPN	GDB:5457725	CORPUS CALLOSUM, AGENESIS OF, WITH NEURONOPATHY
AHO2	GDB:9954535	HEREDITARY OSTEODYSTROPHY-2; AHO2
ANCR	GDB:119678	ANGELMAN SYNDROME
10 B2M	GDB:119028	BETA-2-MICROGLOBULIN; B2M
BBS4	GDB:511199	BARDET-BIEDL SYNDROME, TYPE 4; BBS4
BLM	GDB:135698	BLOOM SYNDROME; BLM
15 CAPN3	GDB:119751	CALPAIN, LARGE POLYPEPTIDE L3; CAPN3 MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2; LGMD2
C DAN1	GDB:9823267	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE I
20 CDAN3	GDB:386192	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE III; CDAN3
CLN6	GDB:4073043	CEROID-LIPOFUSCINOSIS, NEURONAL 6, LATE INFANTILE, VARIANT; CLN6
25 CMH3	GDB:138299	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 3; CMH3
CYP19	GDB:119830	CYTOCHROME P450, SUBFAMILY XIX; CYP19
30 CYP1A1	GDB:120604	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 1; CYP1A1
CYP1A2	GDB:118780	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 2; CYP1A2
35 DYX1	GDB:1391796	DYSLEXIA, SPECIFIC, 1; DYX1

Gene	GDB Accession ID	OMIM Link
EPB42	GDB:127385	HEREDITARY HEMOLYTIC PROTEIN 4.2, ERYTHROCYTIC; EPB42
5 ETFA	GDB:119121	GLUTARICACIDURIA IIA; GA IIA
EYCL3	GDB:4590306	EYE COLOR-3; EYCL3
FAH	GDB:119901	TYROSINEMIA, TYPE I
10 FBN1	GDB:127115	FIBRILLIN-1; FBN1 MARFAN SYNDROME; MFS
FES	GDB:119906	V-FES FELINE SARCOMA VIRAL/V-FPS FUJINAMI AVIAN SARCOMA VIRAL ONCOGENE
HCVS	GDB:119306	CORONAVIRUS 229E SUSCEPTIBILITY; CVS
15 HEXA	GDB:120040	TAY-SACHS DISEASE; TSD
IVD	GDB:119354	ISOVALERICACIDEMIA; IVA
LCS1	GDB:11500552	CHOLESTASIS-LYMPHEDEMA SYNDROME
20 LIPC	GDB:119366	LIPASE, HEPATIC; LIPC
MYO5A	GDB:218824	MYOSIN VA; MYO5A
OCA2	GDB:136820	ALBINISM II
OTSC1	GDB:9860473	OTOSCLEROSIS
PWCR	GDB:120325	PRADER-WILLI SYNDROME
25 RLBP1	GDB:127341	RETINALDEHYDE-BINDING PROTEIN 1,; RLBP1
SLC12A1	GDB:386121	SOLUTE CARRIER FAMILY 12, MEMBER 1; SLC12A1
SPG6	GDB:511201	SPASTIC PARAPLEGIA 6, AUTOSOMAL DOMINANT; SPG6
30 TPM1	GDB:127875	TROPOMYOSIN 1; TPM1
UBE3A	GDB:228487	ANGELMAN SYNDROME UBIQUITIN-PROTEIN LIGASE E3A; UBE3A
35 WMS	GDB:5583902	WEILL-MARCHESANI SYNDROME

Table 17: Genes, Locations and Genetic Disorders on Chromosome 16

Gene	GDB Accession ID	OMIM Link
5 ABCC6	GDB:9315106	PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL DOMINANT; PXE PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL RECESSIVE; PXE
10 ALDOA	GDB:118993	ALDOLASE A, FRUCTOSE-BISPHOSPHATE; ALDOA
15 APRT	GDB:119003	ADENINE PHOSPHORIBOSYLTRANSFERASE; APRT
20 ATP2A1	GDB:119716	ATPase, Ca(2+)-TRANSPORTING, FAST-TWITCH 1; ATP2A1 BRODY MYOPATHY
25 BBS2	GDB:229992	BARDET-BIEDL SYNDROME, TYPE 2; BBS2
30 CARD15	GDB:11026232	SYNOVITIS, GRANULOMATOUS, WITH UVEITIS AND CRANIAL NEUROPATHIES REGIONAL ENTERITIS
35 CATM	GDB:701219	MICROPHTHALMIA-CATARACT
CDH1	GDB:120484	CADHERIN 1; CDH1
CETP	GDB:119773	CHOLESTERYL ESTER TRANSFER PROTEIN, PLASMA; CETP
CHST6	GDB:131407	CORNEAL DYSTROPHY, MACULAR TYPE
CLN3	GDB:120593	CEROID-LIPOFUSCINOSIS, NEURONAL 3, JUVENILE; CLN3
CREBBP	GDB:437159	RUBINSTEIN SYNDROME CREB-BINDING PROTEIN; CREBBP
CTH	GDB:119086	CYSTATHIONINURIA
CTM	GDB:119819	CATARACT, ZONULAR
CYBA	GDB:125238	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-NEGATIVE FORM
CYLD	GDB:701216	EPITHELIOMA, HEREDITARY MULTIPLE BENIGN CYSTIC
DHS	GDB:9958268	XEROCYTOSIS, HEREDITARY
35 DNASE1	GDB:132846	DEOXYRIBONUCLEASE I; DNASE1

)

Gene	GDB Accession ID	OMIM Link
DPEP1	GDB:128059	RENAL DIPEPTIDASE
5 ERCC4	GDB:119113	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 4; ERCC4 XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP F; XPF
FANCA	GDB:701221	FANCONI ANEMIA, COMPLEMENTATION GROUP A; FACA
10 GALNS	GDB:129085	MUCOPOLYSACCHARIDOSIS TYPE IVA
GAN	GDB:9864885	NEUROPATHY, GIANT AXONAL; GAN
HAGH	GDB:119292	HYDROXYACYL GLUTATHIONE HYDROLASE; HAGH
15 HBA1	GDB:119293	HEMOGLOBIN--ALPHA LOCUS-1; HBA1
HBA2	GDB:119294	HEMOGLOBIN--ALPHA LOCUS-2; HBA2
HBHR	GDB:9954541	HEMOGLOBIN H-RELATED MENTAL RETARDATION
HBQ1	GDB:120036	HEMOGLOBIN--THETA-1 LOCUS; HBQ1
20 HBZ	GDB:119302	HEMOGLOBIN--ZETA LOCUS; HBZ
HBZP	GDB:120037	HEMOGLOBIN--ZETA LOCUS; HBZ
HP	GDB:119314	HAPTOGLOBIN; HP
HSD11B2	GDB:409951	CORTISOL 11-BETA-KETOREDUCTASE DEFICIENCY
25 IL4R	GDB:118823	INTERLEUKIN-4 RECEPTOR; IL4R
LIPB	GDB:119365	LIPASE B, LYSOSOMAL ACID; LIPB
MC1R	GDB:135162	MELANOCORTIN-1 RECEPTOR; MC1R
MEFV	GDB:125263	MEDITERRANEAN FEVER, FAMILIAL; MEFV
30 MHC2TA	GDB:6268475	MHC CLASS II TRANSACTIVATOR; MHC2TA
MLYCD	GDB:11500940	MALONYL CoA DECARBOXYLASE DEFICIENCY
35 PHKB	GDB:120286	PHOSPHORYLASE KINASE, BETA SUBUNIT; PHKB

	Gene	GDB Accession ID	OMIM Link
5	PHKG2	GDB:140316	PHOSPHORYLASE KINASE, TESTIS/LIVER, GAMMA 2; PHKG2
	PKD1	GDB:120293	POLYCYSTIC KIDNEY POLYCYSTIC KIDNEY DISEASE 1; PKD1
	PKDTS	GDB:9954545	POLYCYSTIC KIDNEY DISEASE, INFANTILE SEVERE, WITH TUBEROUS SCLEROSIS;
10	PMM2	GDB:438697	CARBOHYDRATE-DEFICIENT GLYCOPROTEIN SYNDROME, TYPE I; CDG1 PHOSPHOMANNOMUTASE 2; PMM2
	PXE	GDB:6053895	PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL DOMINANT; PXE PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL RECESSIVE; PXE
15	SALL1	GDB:4216161	TOWNES-BROCKS SYNDROME; TBS SAL-LIKE 1; SALL1
	SCA4	GDB:250364	SPINOCEREBELLAR ATAXIA 4; SCA4
20	SCNN1B	GDB:434471	SODIUM CHANNEL, NONVOLTAGE-GATED 1 BETA; SCNN1B
	SCNN1G	GDB:568759	SODIUM CHANNEL, NONVOLTAGE-GATED 1 GAMMA; SCNN1G
	TAT	GDB:120398	TYROSINE TRANSAMINASE DEFICIENCY
25	TSC2	GDB:120466	TUBEROUS SCLEROSIS-2; TSC2
	VDI	GDB:119629	DEFECTIVE INTERFERING PARTICLE INDUCTION, CONTROL OF
	WT3	GDB:9958957	WILMS TUMOR, TYPE III; WT3

Table 18: Genes, Locations and Genetic Disorders on Chromosome 17

	Gene	GDB Accession ID	OMIM Link
30	ABR	GDB:119642	ACTIVE BCR-RELATED GENE; ABR
	ACACA	GDB:120534	ACETYL-CoA CARBOXYLASE DEFICIENCY
	ACADVL	GDB:1248185	ACYL-CoA DEHYDROGENASE, VERY-LONG-CHAIN, DEFICIENCY OF
35	ACE	GDB:119840	DIPEPTIDYL CARBOXYPEPTIDASE-1; DCP1

	Gene	GDB Accession ID	OMIM Link
5	ALDH3A2	GDB:1316855	SJOGREN-LARSSON SYNDROME; SLS
	APOH	GDB:118887	APOLIPOPROTEIN H; APOH
	ASPA	GDB:231014	SPONGY DEGENERATION OF CENTRAL NERVOUS SYSTEM
	AXIN2	GDB:9864782	CANCER OF COLON
	BCL5	GDB:125178	LEUKEMIA/LYMPHOMA, CHRONIC B-CELL, 5; BCL5
10	BHD	GDB:11498904	WITH TRICHODISCOMAS AND ACROCHORDONS
	BLMH	GDB:3801467	BLEOMYCIN HYDROLASE
	BRCA1	GDB:126611	BREAST CANCER, TYPE 1; BRCA1
	CACD	GDB:5885801	CHOROIDAL DYSTROPHY, CENTRAL AREOLAR; CACD
	CCA1	GDB:118763	CATARACT, CONGENITAL, CERULEAN TYPE 1; CCA1
15	CCZS	GDB:681973	CATARACT, CONGENITAL ZONULAR, WITH SUTURAL OPACITIES; CCZS
	CHRNB1	GDB:120587	CHOLINERGIC RECEPTOR, NICOTINIC, BETA POLYPEPTIDE 1; CHRNB1
	CHRNE	GDB:132246	CHOLINERGIC RECEPTOR, NICOTINIC, EPSILON POLYPEPTIDE; CHRNE
	CMT1A	GDB:119785	CHARCOT-MARIE-TOOTH DISEASE, TYPE 1A; CMT1A NEUROPATHY, HEREDITARY, WITH LIABILITY TO PRESSURE PALSIES; HNPP
	COL1A1	GDB:119061	COLLAGEN, TYPE I, ALPHA-1 CHAIN; COL1A1 OSTEOGENESIS IMPERFECTA TYPE I OSTEOGENESIS IMPERFECTA TYPE IV; OI4
20	CORDS	GDB:568473	CONE-ROD DYSTROPHY-5; CORD5
	CTNS	GDB:700761	CYSTINOSIS, EARLY-ONSET OR INFANTILE NEPHROPATHIC TYPE
	EPX	GDB:377700	EOSINOPHIL PEROXIDASE; EPX

Gene	GDB Accession ID	OMIM Link
ERBB2	GDB:120613	V-ERB-B2 AVIAN ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE HOMOLOG 2; ERBB2
5 G6PC	GDB:231927	GLYCOGEN STORAGE DISEASE I; GSD-I
GAA	GDB:119965	GLYCOGEN STORAGE DISEASE II
GALK1	GDB:119246	GALACTOKINASE DEFICIENCY
10 GCGR	GDB:304516	GLUCAGON RECEPTOR; GCGR
GFAP	GDB:118799	GLIAL FIBRILLARY ACIDIC PROTEIN; GFAP ALEXANDER DISEASE
GH1	GDB:119982	GROWTH HORMONE 1; GH1
GH2	GDB:119983	GROWTH HORMONE 2; GH2
15 GP1BA	GDB:118806	GIANT PLATELET SYNDROME
GPSC	GDB:9954564	FAMILIAL PROGRESSIVE SUBCORTICAL
GUCY2D	GDB:136012	AMAUROSIS CONGENITA OF LEBER 1 GUANYLATE CYCLASE 2D, MEMBRANE; GUC2D CONE-ROD DYSTROPHY-6; CORD6
20 ITGA2B	GDB:120012	THROMBASTHENIA OF GLANZMANN AND NAEGELEI
ITGB3	GDB:120013	INTEGRIN, BETA-3; ITGB3
ITGB4	GDB:128028	INTEGRIN, BETA-4; ITGB4
25 KRT10	GDB:118828	KERATIN 10; KRT10
KRT12	GDB:5583953	CORNEAL DYSTROPHY, JUVENILE EPITHELIAL, OF MEESMANN KERATIN 12; KRT12
KRT13	GDB:120740	KERATIN 13; KRT13
30 KRT14	GDB:132145	KERATIN 14; KRT14 GLUTATHIONE SYNTHETASE; GSS
KRT14L1	GDB:120121	KERATIN 14; KRT14
KRT14L2	GDB:120122	KERATIN 14; KRT14
KRT14L3	GDB:120123	KERATIN 14; KRT14
35 KRT16	GDB:136207	KERATIN 16; KRT16

Gene	GDB Accession ID	OMIM Link
KRT16L1	GDB:120125	KERATIN 16; KRT16
KRT16L2	GDB:120126	KERATIN 16; KRT16
5	KRT17	KERATIN 17; KRT17 PACHYONYCHIA CONGENITA, JACKSON-LAWLER TYPE
	KRT9	HYPERKERATOSIS, LOCALIZED EPIDERMOLYTIC
	MAPT	MICROTUBULE-ASSOCIATED PROTEIN TAU; MAPT PALLIDOPONTONIGRAL DEGENERATION; PPND DISINHIBITION-DEMENTIA-PARKINSONI SM-AMYOTROPHY COMPLEX; DDPAC
	MDB	MEDULLOBLASTOMA; MDB
10	MDCR	MILLER-DIEKER LISSENCEPHALY SYNDROME; MDLS PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
	MGI	MYASTHENIA GRAVIS, FAMILIAL INFANTILE; FIMG
	MHS2	MALIGNANT HYPERHERMIA SUSCEPTIBILITY-2; MHS2
	MKS1	MECKEL SYNDROME; MKS
15	MPO	MYELOPEROXIDASE DEFICIENCY
	MUL	MULIBREY NANISM; MUL
	MYO15A	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 3; DFNB3
	NAGLU	MUCOPOLYSACCHARIDOSIS TYPE III B
20	NAPB	NEURITIS WITH BRACHIAL PREDILECTION; NAPB
	NF1	NEUROFIBROMATOSIS, TYPE I; NF1
	NME1	NON-METASTATIC CELLS 1, PROTEIN EXPRESSED IN; NME1
	P4HB	PROLYL-4-HYDROXYLASE, BETA POLYPEPTIDE; PHDB; PROHB
25		
30		

	Gene	GDB Accession ID	OMIM Link
5	PAFAH1B1	GDB:677430	MILLER-DIEKER LISSENCEPHALY SYNDROME; MDLS PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
	PECAM1	GDB:696372	PLATELET-ENDOTHELIAL CELL ADHESION MOLECULE; PECAM1
10	PEX12	GDB:6155804	ZELLWEGER SYNDROME; ZS PEROXIN-12; PEX12
	PHB	GDB:126600	PROHIBITIN; PHB
15	PMP22	GDB:134190	CHARCOT-MARIE-TOOTH DISEASE, TYPE 1A; CMT1A HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS PERIPHERAL MYELIN PROTEIN 22; PMP22
	PRKAR1A	GDB:120313	MYXOMA, SPOTTY PIGMENTATION, AND ENDOCRINE OVERACTIVITY PROTEIN KINASE, cAMP-DEPENDENT, REGULATORY, TYPE I, ALPHA; PRKAR1A
20	PRKCA	GDB:128015	PROTEIN KINASE C, ALPHA; PRKCA
	PRKWNK4	GDB:9954566	PSEUDOHYPOALDOSTERONISM TYPE II, LOCUS B; PHA2B
25	PRP8	GDB:9957697	RETINITIS PIGMENTOSA-13; RP13
	PRPF8	GDB:392647	RETINITIS PIGMENTOSA-13; RP13
30	PTLAH	GDB:9957342	APLASIA OR HYPOPLASIA
	RARA	GDB:120337	RETINOIC ACID RECEPTOR, ALPHA; RARA
35	RCV1	GDB:135477	RECOVERIN; RCV1
	RMSA1	GDB:304519	REGULATOR OF MITOTIC SPINDLE ASSEMBLY 1; RMSA1
	RP17	GDB:683199	RETINITIS PIGMENTOSA-17; RP17
	RSS	GDB:439249	RUSSELL-SILVER SYNDROME; RSS
	SCN4A	GDB:125181	PERIODIC PARALYSIS II
	SERPINF2	GDB:120301	PLASMIN INHIBITOR DEFICIENCY
	SGCA	GDB:384077	ADHALIN; ADL

)

Gene	GDB Accession ID	OMIM Link
SGSH	GDB:1319101	MUCOPOLYSACCHARIDOSIS TYPE IIIA
SHBG	GDB:125280	SEX HORMONE BINDING GLOBULIN; SHBG
5 SLC2A4	GDB:119997	SOLUTE CARRIER FAMILY 2, MEMBER 4; SLC2A4
10 SLC4A1	GDB:119874	SOLUTE CARRIER FAMILY 4, ANION EXCHANGER, MEMBER 1; SLC4A1 BLOOD GROUP--DIEGO SYSTEM; DI BLOOD GROUP--WRIGHT ANTIGEN; Wr ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE HEREDITARY HEMOLYTIC
15 SLC6A4	GDB:134713	SOLUTE CARRIER FAMILY 6, MEMBER 4; SLC6A4
SMCR	GDB:120379	SMITH-MAGENIS SYNDROME; SMS
15 SOST	GDB:10450629	SCLEROSTESIS
SOX9	GDB:134730	DYSPLASIA
20 SSTR2	GDB:134186	SOMATOSTATIN RECEPTOR-2; SSTR2
SYM1	GDB:512174	SYMPHALANGISM, PROXIMAL; SYM1
20 SYNS1	GDB:9862343	SYNOSTOSES, MULTIPLE, WITH BRACHYDACTYLY
TCF2	GDB:125298	TRANSCRIPTION FACTOR-2, HEPATIC; TCF2
25 THRA	GDB:120730	THYROID HORMONE RECEPTOR, ALPHA 1; THRA
TIMP2	GDB:132612	TISSUE INHIBITOR OF METALLOPROTEINASE-2; TIMP2
TOC	GDB:451978	TYLOSIS WITH ESOPHAGEAL CANCER; TOC
30 TOP2A	GDB:118884	TOPOISOMERASE (DNA) II, ALPHA; TOP2A
TP53	GDB:120445	CANCER, HEPATOCELLULAR LI-FRAUMENI SYNDROME; LFS TUMOR PROTEIN p53; TP53 CARCINOMA
35 VBCH	GDB:9954554	HYPEROSTOSIS CORTICALIS GENERALISATA

Table 19: Genes, Locations and Genetic Disorders on Chromosome 18

Gene	GDB Accession ID	OMIM Link
ATP8B1	GDB:453352	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATIC 1; PFIC1 INTRAHEPATIC CHOLESTASIS FAMILIAL INTRAHEPATIC CHOLESTASIS-1; FIC1
BCL2	GDB:119031	B-CELL CLL/LYMPHOMA 2; BCL2
CNSN	GDB:9954580	CARNOSINEMIA
CORD1	GDB:118773	CONE-ROD DYSTROPHY-1; CORD1
CYB5	GDB:125236	METHEMOGLOBINEMIA DUE TO DEFICIENCY OF CYTOCHROME b5
DCC	GDB:119838	DELETED IN COLORECTAL CARCINOMA; DCC
F5F8D	GDB:6919858	FACTOR V AND FACTOR VIII, COMBINED DEFICIENCY OF; F5F8D
FECH	GDB:127282	PROTOPORPHYRIA, ERYTHROPOIETIC
FEO	GDB:4378120	POLYOSTOTIC OSTEOLYTIC DYSPLASIA, HEREDITARY EXPANSILE; HEPOD
LAMA3	GDB:251818	LAMININ, ALPHA 3; LAMA3
LCFS2	GDB:9954578	CANCER
MADH4	GDB:4642788	POLYPOSIS, JUVENILE INTESTINAL MOTHERS AGAINST DECAPENTAPLEGIC, DROSOPHILA, HOMOLOG OF, 4; MADH4
MAFD1	GDB:120163	MANIC-DEPRESSIVE PSYCHOSIS, AUTOSOMAL
MC2R	GDB:135163	ADRENAL UNRESPONSIVENESS TO ACTH
MCL	GDB:9954574	LEIOMYOMATA, HEREDITARY MULTIPLE, OF SKIN
MYP2	GDB:9862232	MYOPIA
NPC1	GDB:138178	NIEMANN-PICK DISEASE, TYPE C1; NPC1
SPPK	GDB:606444	PALMOPLANTARIS STRIATA
TGFBRE	GDB:250852	TRANSFORMING GROWTH FACTOR, BETA 1 RESPONSE ELEMENT
TGIF	GDB:9787150	HOLOPROSENCEPHALY, TYPE 4; HPE4
TTR	GDB:119471	TRANSTHYRETIN; TTR

Table 20: Genes, Locations and Genetic Disorders on Chromosome 19

Gene	GDB Accession ID	OMIM Link
AD2	GDB:118748	ALZHEIMER DISEASE-2; AD2
AMH	GDB:118996	PERSISTENT MULLERIAN DUCT SYNDROME, TYPES I AND II; PMDS ANTI-MULLERIAN HORMONE; AMH
APOC2	GDB:119689	APOLIPOPROTEIN C-II DEFICIENCY, TYPE I HYPERLIPOPROTEINEMIA DUE TO
APOE	GDB:119691	APOLIPOPROTEIN E; APOE
ATHS	GDB:128803	LIPOPROTEIN PHENOTYPE; ALP
BAX	GDB:228082	BCL2-ASSOCIATED X PROTEIN; BAX
BCKDHA	GDB:119723	MAPLE SYRUP URINE DISEASE
BCL3	GDB:120561	B-CELL LEUKEMIA/LYMPHOMA-3; BCL3
BFIC	GDB:9954584	BENIGN FAMILIAL INFANTILE CONVULSIONS
C3	GDB:119044	COMPLEMENT COMPONENT-3; C3
CACNA1A	GDB:126432	ATAxia, PERIODIC VESTIBULOCEREBELLAR HEMIPLEGIC MIGRAINE, FAMILIAL; MHP SPINOCEREBELLAR ATAXIA 6; SCA6 CALCIUM CHANNEL, VOLTAGE-DEPENDENT, P/Q TYPE, ALPHA 1A SUBUNIT; CACNA1A
CCO	GDB:119755	CENTRAL CORE DISEASE OF MUSCLE
CEACAM5	GDB:119054	CARCINOEMBRYONIC ANTIGEN; CEA
COMP	GDB:344263	EPiphyseal dysplasia, multiple; MED PSEUDOACHONDROPLASTIC DYSPLASIA CARTILAGE OLIGOMERIC MATRIX PROTEIN; COMP
CRX	GDB:333932	CONE-ROD DYSTROPHY-2; CORD2 AMAUROSIS CONGENITA OF LEBER I CONE-ROD HOMEOPROTEIN
DBA	GDB:9600353	ANEMIA, CONGENITAL HYPOPLASTIC, OF BLACKFAN AND DIAMOND

Gene	GDB Accession ID	OMIM Link
DDU	GDB:10796026	URTICARIA; DDU
5 DFNA4	GDB:606540	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 4; DFNA4
DLL3	GDB:9959026	VERTEBRAL ANOMALIES
DMPK	GDB:119097	DYSTROPHIA MYOTONICA; DM
10 DMWD	GDB:7178354	DYSTROPHIA MYOTONICA; DM
DPD1	GDB:10796170	ENGELMANN DISEASE
E11S	GDB:119101	ECHO 11 SENSITIVITY; E11S
ELA2	GDB:118792	ELASTASE-2; ELA2 NEUTROOPENIA, CYCLIC
15 EPOR	GDB:125242	ERYTHROPOIETIN RECEPTOR; EPOR
ERCC2	GDB:119112	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 2; ERCC2 XERODERMA PIGMENTOSUM IV; XP4
20 ETFB	GDB:119887	ELECTRON TRANSFER FLAVOPROTEIN, BETA POLYPEPTIDE; ETFB
EXT3	GDB:383780	EXOSTOSES, MULTIPLE, TYPE III; EXT3
EYCL1	GDB:119269	EYE COLOR-1; EYCL1
FTL	GDB:119234	FERRITIN LIGHT CHAIN; FTL
25 FUT1	GDB:120618	FUCOSYLTRANSFERASE-1; FUT1
FUT2	GDB:120619	FUCOSYLTRANSFERASE-2; FUT2
FUT6	GDB:135180	FUCOSYLTRANSFERASE-6; FUT6
GAMT	GDB:1313736	GUANIDINOACETATE METHYLTRANSFERASE; GAMT
30 GCDH	GDB:136004	GLUTARICACIDEMIA I
GPI	GDB:120015	GLUCOSEPHOSPHATE ISOMERASE; GPI
GUSM	GDB:119291	GLUCURONIDASE, MOUSE, MODIFIER OF; GUSM
35 HB1	GDB:9954586	BUNDLE BRANCH BLOCK
HCL1	GDB:119304	HAIR COLOR-1; HCL1

	Gene	GDB Accession ID	OMIM Link
	HHC2	GDB:249836	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL, TYPE II; HHC2
5	HHC3	GDB:9955121	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL, TYPE III; HHC3
	ICAM3	GDB:136236	INTERCELLULAR ADHESION MOLECULE-3; ICAM3
	INSR	GDB:119352	INSULIN RECEPTOR; INSR
10	JAK3	GDB:376460	JANUS KINASE 3 JAK3
	KLK3	GDB:119695	ANTIGEN, PROSTATE-SPECIFIC; APS
	LDLR	GDB:119362	HYPERCHOLESTEROLEMIA, FAMILIAL; FHC
15	LHB	GDB:119364	LUTEINIZING HORMONE, BETA POLYPEPTIDE; LHB
	LIG1	GDB:127274	LIGASE I, DNA, ATP-DEPENDENT; LIG1
	LOH19CR1	GDB:9837482	ANEMIA, CONGENITAL HYPOPLASTIC, OF BLACKFAN AND DIAMOND
	LYL1	GDB:120158	LEUKEMIA, LYMPHOID, 1; LYL1
20	MAN2B1	GDB:119376	MANNOSIDOSIS, ALPHA B, LYSOSOMAL
	MCOLN1	GDB:10013974	MUCOLIPIDOSIS IV
	MDRV	GDB:6306714	MUSCULAR DYSTROPHY, AUTOSOMAL DOMINANT, WITH RIMMED VACUOLES; MDRV
25	MLLT1	GDB:136791	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 1; MLLT1
	NOTCH3	GDB:361163	DEMENTIA, HEREDITARY MULTI-INFARCT TYPE NOTCH, DROSOPHILA, HOMOLOG OF, 3; NOTCH3
30	NPHS1	GDB:342105	NEPHROSIS 1, CONGENITAL, FINNISH TYPE; NPHS1
	OFC3	GDB:128060	OROFACIAL CLEFT-3; OFC3
35	OPA3	GDB:9954590	OPTIC ATROPHY, INFANTILE, WITH CHOREA AND SPASTIC PARAPLEGIA

Gene	GDB Accession ID	OMIM Link
PEPD	GDB:120273	PEPTIDASE D; PEPD
PRPF31	GDB:333911	RETINITIS PIGMENTOSA 11; RP11
PRTN3	GDB:126876	PROTEINASE 3; PRTN3; PR3
PRX	GDB:11501256	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS
PSG1	GDB:120321	PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN 1; PSG1
PVR	GDB:120324	POLIOVIRUS SUSCEPTIBILITY, OR SENSITIVITY; PVS
RYR1	GDB:120359	CENTRAL CORE DISEASE OF MUSCLE HYPERTERMIA OF ANESTHESIA RYANODINE RECEPTOR-1; RYR1
SLC5A5	GDB:5892184	SOLUTE CARRIER FAMILY 5, MEMBER 5; SLC5A5
SLC7A9	GDB:9958852	CYSTINURIA, TYPE III; CSNU3
STK11	GDB:9732383	PEUTZ-JEGHERS SYNDROME SERINE/THREONINE PROTEIN KINASE 11; STK11
TBXA2R	GDB:127517	THROMBOXANE A2 RECEPTOR, PLATELET; TBXA2R
TGFB1	GDB:120729	ENGELMANN DISEASE TRANSFORMING GROWTH FACTOR, BETA-1; TGFB1
TNNI3	GDB:125309	TROPONIN I, CARDIAC; TNNI3
TYROBP	GDB:9954457	POLYCYSTIC LIPOMEMBRANOUS OSTEODYSPLOASIA WITH SCLEROSING LEUKOENCEPHALOPATHY

Table 21: Genes, Locations and Genetic Disorders on Chromosome 20

Gene	GDB Accession ID	OMIM Link
ADA	GDB:119649	ADENOSINE DEAMINASE; ADA
AHCY	GDB:118983	S-ADENOSYLHOMOCYSTEINE HYDROLASE; AHCY
AVP	GDB:119009	DIABETES INSIPIDUS, NEUROHYPOPHYSEAL TYPE ARGININE VASOPRESSIN; AVP

	Gene	GDB Accession ID	OMIM Link
	CDAN2	GDB:9823270	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE II
5	CDMP1	GDB:438940	CHONDRODYSPLASIA, GREBE TYPE; CARTILAGE-DERIVED MORPHOGENETIC PROTEIN 1
	CHED1	GDB:3837719	CORNEAL DYSTROPHY, CONGENITAL; ENDOTHELIAL; CHED
10	CHRNA4	GDB:128169	CHOLINERGIC RECEPTOR, NEURONAL; NICOTINIC, ALPHA POLYPEPTIDE 4; CHRNA4 EPILEPSY, BENIGN NEONATAL; EBN1
	CST3	GDB:119817	AMYLOIDOSIS VI
15	EDN3	GDB:119862	ENDOTHELIN-3; EDN3 WAARDENBURG-SHAH SYNDROME
	EEGV1	GDB:127525	ELECTROENCEPHALOGRAM, LOW-VOLTAGE
	FTLL1	GDB:119235	FERRITIN LIGHT CHAIN; FTL
20	GNAS	GDB:120628	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-STIMULATING POLYPEPTIDE;
	GSS	GDB:637022	GLUTATHIONE SYNTHETASE DEFICIENCY OF ERYTHROCYTES, HEMOLYTIC ANEMIA, PYROGLUTAMICACIDURIA, HNF4AGDB:393281DIABETES MELLITUS, AUTOSOMAL DOMINANT TRANSCRIPTION FACTOR 14, HEPATIC NUCLEAR FACTOR; TCF14
25	JAG1	GDB:6175920	CHOLESTASIS WITH PERIPHERAL PULMONARY STENOSIS JAGGED 1; JAG1
	KCNQ2	GDB:9787229	EPILEPSY, BENIGN NEONATAL; EBN1 POTASSIUM CHANNEL, VOLTAGE-GATED, SUBFAMILY Q, MEMBER 2
30	MKKS	GDB:9860197	HYDROMETROCOLPOS SYNDROME
	NBIA1	GDB:4252819	HALLERVORDEN-SPATZ DISEASE
35	PCK1	GDB:125349	PHOSPHOENOLPYRUVATE CARBOXYKINASE 1, SOLUBLE; PCK1

Gene	GDB Accession ID	OMIM Link
PI3	GDB:203940	PROTEINASE INHIBITOR 3; PI3
PPGB	GDB:119507	NEURAMINIDASE DEFICIENCY WITH BETA-GALACTOSIDASE DEFICIENCY
PPMD	GDB:702144	CORNEAL DYSTROPHY, HEREDITARY POLYMORPHOUS POSTERIOR; PPCD
PRNP	GDB:120720	GERSTMANN-STRAUSSLER DISEASE; GSD PRION PROTEIN; PRNP
THBD	GDB:119613	THROMBOMODULIN; THBD
TOP1	GDB:120444	TOPOISOMERASE (DNA) I; TOP1

Table 22: Genes, Locations and Genetic Disorders on Chromosome 21

Gene	GDB Accession ID	OMIM Link
AIRE	GDB:567198	AUTOIMMUNE POLYENDOCRINOPATHY-CANDIDIASIS-ECTODERMAL DYSTROPHY; APECED
APP	GDB:119692	ALZHEIMER DISEASE; AD AMYLOID BETA A4 PRECURSOR PROTEIN; APP
CBS	GDB:119754	HOMOCYSTINURIA
COL6A1	GDB:119065	COLLAGEN, TYPE VI, ALPHA-1 CHAIN; COL6A1 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
COL6A2	GDB:119793	COLLAGEN, TYPE VI, ALPHA-2 CHAIN; COL6A2 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
CSTB	GDB:5215249	MYOCLONUS EPILEPSY OF UNVERRICHT AND LUNDBORG CYSTATIN B; CSTB
DCR	GDB:125354	TRISOMY 21
DSCR1	GDB:731000	TRISOMY 21
FPDMM	GDB:9954610	CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 2; CBFA2 PLATELET DISORDER, FAMILIAL, WITH ASSOCIATED MYELOID MALIGNANCY

Gene	GDB Accession ID	OMIM Link
HLCS	GDB:392648	MULTIPLE CARBOXYLASE DEFICIENCY, BIOTIN-RESPONSIVE; MCD
HPE1	GDB:136065	HOLOPROSENCEPHALY, FAMILIAL ALOBAR
ITGB2	GDB:120574	INTEGRIN BETA-2; ITGB2
KCNE1	GDB:127909	POTASSIUM VOLTAGE-GATED CHANNEL, ISK-RELATED SUBFAMILY, MEMBER 1;
KNO	GDB:4073044	KNOBLOCH SYNDROME; KNO
PRSS7	GDB:384083	ENTEROKINASE DEFICIENCY
RUNX1	GDB:128313	CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 2; CBFA2 PLATELET DISORDER, FAMILIAL, WITH ASSOCIATED MYELOID MALIGNANCY
SOD1	GDB:119596	AMYOTROPHIC LATERAL SCLEROSIS SUPEROXIDE DISMUTASE-1; SOD1 MUSCULAR ATROPHY, PROGRESSIVE, WITH AMYOTROPHIC LATERAL SCLEROSIS
TAM	GDB:9958709	MYELOPROLIFERATIVE SYNDROME, TRANSIENT

Table 23: Genes, Locations and Genetic Disorders on Chromosome 22

Gene	GDB Accession ID	OMIM Link
ADSL	GDB:119655	ADENYLOSUCCINATE LYASE; ADSL
ARSA	GDB:119007	METACHROMATIC LEUKODYSTROPHY, LATE-INFANTILE
BCR	GDB:120562	BREAKPOINT CLUSTER REGION; BCR
COCR	GDB:119772	CAT EYE SYNDROME; CES
CHEK2	GDB:9958730	LI-FRAUMENI SYNDROME; LFS OSTEOGENIC SARCOMA
COMT	GDB:119795	CATECHOL-O-METHYLTRANSFERASE ; COMT

Gene	GDB Accession ID	OMIM Link
CRYBB2	GDB:119075	CRYSTALLIN, BETA B2; CRYBB2 CATARACT, CONGENITAL, CERULEAN TYPE, 2; CCA2
5 CSF2RB	GDB:126838	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR, BETA SUBUNIT;
10 CTHM	GDB:439247	HEART MALFORMATIONS; CTHM
10 CYP2D6	GDB:132127	CYTOCHROME P450, SUBFAMILY IID; CYP2D
15 CYP2D@	GDB:119832	CYTOCHROME P450, SUBFAMILY IID; CYP2D
DGCR	GDB:119843	DIGEORGE SYNDROME; DGS
15 DIA1	GDB:119848	METHEMOGLOBINEMIA DUE TO DEFICIENCY OF METHEMOGLOBIN REDUCTASE
EWSR1	GDB:135984	EWING SARCOMA; EWS
GGT1	GDB:120623	GLUTATHIONURIA
20 MGCR	GDB:120180	MENINGIOMA; MGM
20 MN1	GDB:580528	MENINGIOMA; MGM
NAGA	GDB:119445	ALPHA-GALACTOSIDASE B; GALB
NF2	GDB:120232	NEUROFIBROMATOSIS, TYPE II; NF2
25 OGS2	GDB:9954619	HYPERTELORISM WITH ESOPHAGEAL ABNORMALITY AND HYPOSPADIAS
PDGFB	GDB:120709	V-SIS PLATELET-DERIVED GROWTH FACTOR BETA POLYPEPTIDE; PDGFB
30 PPARA	GDB:202877	PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR, ALPHA; PPARA
30 PRODH	GDB:5215168	HYPERPROLINEMIA, TYPE I
SCO2	GDB:9958568	CYTOCHROME c OXIDASE DEFICIENCY
35 SCZD4	GDB:1387047	SCHIZOPHRENIA DISORDER-4; SCZD4
35 SERPIND1	GDB:120038	HEPARIN COFACTOR II; HCF2

Gene	GDB Accession ID	OMIM Link
5	SLC5A1	SOLUTE CARRIER FAMILY 5, MEMBER 1; SLC5A1
	SOX10	SRY-BOX 10; SOX10
	TCN2	TRANSCOBALAMIN II DEFICIENCY
	TIMP3	TISSUE INHIBITOR OF METALLOPROTEINASE-3; TIMP3
	VCF	VELOCARDIOFACIAL SYNDROME

10

Table 24: Genes, Locations and Genetic Disorders on Chromosome X

Gene	GDB Accession ID	OMIM Link
15	ABCD1	ADRENOLEUKODYSTROPHY; ALD
	ACTL1	ACTIN-LIKE SEQUENCE-1; ACTL1
	ADFN	ALBINISM-DEAFNESS SYNDROME; ADFN; ALDS
	AGMX2	AGAMMAGLOBULINEMIA, X-LINKED, TYPE 2; AGMX2; XLA2
	AHDS	MENTAL RETARDATION, X-LINKED, WITH HYPOTONIA
20	AIC	CORPUS CALLOSUM, AGENESIS OF, WITH CHORIORETINAL ABNORMALITY
	AIED	ALBINISM, OCULAR, TYPE 2; OA2
	AIH3	AMELOGENESIS IMPERFECTA-3, HYPOPLASTIC TYPE; AIH3
	ALAS2	ANEMIA, HYPOCHROMIC
	AMCD	ARTHROGRYPOSIS MULTIPLEX CONGENITA, DISTAL
25	AMELX	AMELOGENESIS IMPERFECTA-1, HYPOPLASTIC TYPE; AIH1
	ANOP1	CLINICAL; ANOP1
	AR	ANDROGEN INSENSITIVITY SYNDROME; AIS ANDROGEN RECEPTOR; AR
	ARAF1	V-RAF MURINE SARCOMA 3611 VIRAL ONCOGENE HOMOLOG 1; ARAF1

Gene	GDB Accession ID	OMIM Link
ARSC2	GDB:119702	ARYLSULFATASE C, f FORM; ARSC2
5	ARSE	CHONDRODYSPLASIA PUNCTATA 1, X-LINKED RECESSIVE; CDPX1
	ARTS	FATAL X-LINKED, WITH DEAFNESS AND LOSS OF VISION
	ASAT	SIDEROBLASTIC, AND SPINOCEREBELLAR ATAXIA; ASAT
	ASSP5	CITRULLINEMIA
10	ATP7A	ATPase, Cu(2+)-TRANSPORTING, ALPHA POLYPEPTIDE; ATP7A MENKES SYNDROME
	ATRX	ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, X-LINKED; ATRX ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, NONDELETION TYPE
	AVPR2	DIABETES INSIPIDUS, NEPHROGENIC
15	BFLS	BORJESON SYNDROME; BORJ
	BGN	BIGLYCAN; BGN
	BTK	BRUTON AGAMMAGLOBULINEMIA TYROSINE KINASE; BTK
	BZX	BAZEX SYNDROME; BZX
20	C1HR	TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR 2A; TAF2A
	CACNA1F	NIGHTBLINDNESS, CONGENITAL STATIONARY, X-LINKED, TYPE 2; CSNB2 CALCIUM CHANNEL, VOLTAGE-DEPENDENT, ALPHA 1F SUBUNIT; CACNA1F
	CALB3	CALBINDIN 3; CALB3
25	CBBM	COLORBLINDNESS, BLUE-MONO-CONE-MONOCHROMATIC TYPE; CBBM
	CCT	CATARACT, CONGENITAL TOTAL, WITH POSTERIOR SUTURAL OPACITIES IN HETEROZYGOTES;
30		
35		

Gene	GDB Accession ID	OMIM Link
CDR1	GDB:119053	CEREBELLAR DEGENERATION-RELATED AUTOANTIGEN-1; CDR1; CDR34
5 CFNS	GDB:9579470	CRANIOFRONTONASAL SYNDROME; CFNS
CGF1	GDB:6275867	COGNITION
CHM	GDB:120400	CHOROIDEREMIA; CHM
10 CHR39C	GDB:119779	CHOLESTEROL REPRESSIBLE PROTEIN 39C; CHR39C
CIDX	GDB:127736	SEVERE COMBINED IMMUNODEFICIENCY DISEASE, X-LINKED, 2; SCIDX2
CLA2	GDB:119782	CEREBELLAR ATAXIA, X-LINKED; CLA2
15 CLCN5	GDB:270667	CHLORIDE CHANNEL 5; CLCN5 FANCONI SYNDROME, RENAL, WITH NEPHROCALCINOSIS AND RENAL STONES NEPHROLITHIASIS, X-LINKED RECESSIVE, WITH RENAL FAILURE; XRN
20 CLS	GDB:119784	RIBOSOMAL PROTEIN S6 KINASE, 90 KD, POLYPEPTIDE 3; RPS6KA3 COFFIN-LOWRY SYNDROME; CLS
CMTX2	GDB:128311	CHARCOT-MARIE-TOOTH NEUROPATHY, X-LINKED RECESSIVE, 2; CMTX2
CMTX3	GDB:128151	CHARCOT-MARIE-TOOTH NEUROPATHY, X-LINKED RECESSIVE, 3; CMTX3
25 CND	GDB:9954627	DERMOIDS OF CORNEA; CND
COD1	GDB:119787	CONE DYSTROPHY, X-LINKED, 1; COD1
COD2	GDB:6520166	CONE DYSTROPHY, X-LINKED, 2; COD2
30 COL4A5	GDB:120596	COLLAGEN, TYPE IV, ALPHA-5 CHAIN; COL4A5 LEIOMYOMATOSIS, ESOPHAGEAL AND VULVAL, WITH NEPHROPATHY
COL4A6	GDB:222775	COLLAGEN, TYPE IV, ALPHA-6 CHAIN; COL4A6 LEIOMYOMATOSIS, ESOPHAGEAL AND VULVAL, WITH NEPHROPATHY
35 CPX	GDB:120598	CLEFT PALATE, X-LINKED; CPX

	Gene	GDB Accession ID	OMIM Link
5	CVD1	GDB:9954659	CARDIAC VALVULAR DYSPLASIA, X-LINKED
	CYBB	GDB:120513	GRANULOMATOUS DISEASE, CHRONIC; CGD
	DCX	GDB:9823272	LISSENCEPHALY, X-LINKED
	DFN2	GDB:119091	DEAFNESS, X-LINKED 2, PERCEPTIVE CONGENITAL; DFN2
10	DFN4	GDB:433255	DEAFNESS, X-LINKED 4, CONGENITAL SENSORINEURAL; DFN4
	DFN6	GDB:1320698	DEAFNESS, X-LINKED, 6, PROGRESSIVE; DFN6
	DHOF	GDB:119847	FOCAL DERMAL HYPOPLASIA; DHOF
15	DIAPH2	GDB:9835484	DIAPHANOUS, DROSOPHILA, HOMOLOG OF, 2 DKC1GDB:119096 DYSKERATOSIS CONGENITA; DKC
	DMD	GDB:119850	MUSCULAR DYSTROPHY, PSEUDOHYPERTROPHIC PROGRESSIVE, DUCHENNE AND BECKER
20	DSS	GDB:433750	DOSAGE-SENSITIVE SEX REVERSAL; DSS
	DYT3	GDB:118789	TORSION DYSTONIA-3, X-LINKED TYPE; DYT3
25	EBM	GDB:119102	BULLOUS DYSTROPHY, HEREDITARY MACULAR TYPE
	EBP	GDB:125212	CHONDRODYSPLASIA PUNCTATA, X-LINKED DOMINANT; CDPX2; CDPXD; CPXD
	ED1	GDB:119859	ECTODERMAL DYSPLASIA, ANHIDROTIC; EDA
30	ELK1	GDB:119867	ELK1, MEMBER OF ETS ONCOGENE FAMILY; ELK1
	EMD	GDB:119108	MUSCULAR DYSTROPHY, TARDIVE, DREIFUSS-EMERY TYPE, WITH CONTRACTURES
35	EVR2	GDB:136068	EXUDATIVE VITREORETINOPATHY, FAMILIAL, X-LINKED RECESSIVE; EVR2
	F8C	GDB:119124	HEMOPHILIA A

	Gene	GDB Accession ID	OMIM Link
5	F9	GDB:119900	HEMOPHILIA B; HEMB
	FCP1	GDB:347490	F-CELL PRODUCTION, X-LINKED; FCPX
	FDPSL5	GDB:119922	SYNTHETASE-5; FPSL5
	FGD1	GDB:119131	SYNDROME FACIOGENITAL DYSPLASIA; FGDY
	FGS1	GDB:9836950	FG SYNDROME
10	FMR1	GDB:129038	FRAGILE SITE MENTAL RETARDATION-1; FMR1
	FMR2	GDB:141566	FRAGILE SITE, FOLIC ACID TYPE, RARE, FRA(X)(q28); FRAXE
	G6PD	GDB:120621	GLUCOSE-6-PHOSPHATE DEHYDROGENASE; G6PD
15	GABRA3	GDB:119968	GAMMA-AMINOBUTYRIC ACID RECEPTOR, ALPHA-3; GABRA3
	GATA1	GDB:125373	GATA-BINDING PROTEIN 1; GATA1
	GDI1	GDB:1347097	GDP DISSOCIATION INHIBITOR 1; GDI1 MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 3; MRX3
20	GDXY	GDB:9954629	DYSGENESIS, XY FEMALE TYPE; GDXY
	GJB1	GDB:125246	CHARCOT-MARIE-TOOTH PERONEAL MUSCULAR ATROPHY, X-LINKED; CMTX1 GAP JUNCTION PROTEIN, BETA-1, 32 KD; GJB1
	GK	GDB:119271	HYPERGLYCEMIA
25	GLA	GDB:119272	ANGIokeratoma, DIFFUSE
	GPC3	GDB:3770726	GLYCAN-3; GPC3 SIMPSON DYSMORPHIA SYNDROME; SDYS
	GRPR	GDB:128035	GASTRIN-RELEASING PEPTIDE RECEPTOR; GRPR
30	GTD	GDB:9954635	GONADOTROPIN DEFICIENCY; GTD
	GUST	GDB:9954655	MENTAL RETARDATION WITH OPTIC ATROPHY, DEAFNESS, AND SEIZURES
	HMS1	GDB:251827	1; HMS1

Gene	GDB Accession ID	OMIM Link
HPRT1	GDB:119317	HYPOXANTHINE GUANINE PHOSPHORIBOSYLTRANSFERASE 1; HPRT1
HPT	GDB:119322	HYPOPARTHYROIDISM, X-LINKED; HYPX
HTC2	GDB:700980	HYPERTRICHOSIS, CONGENITAL GENERALIZED; CGH; HCG
HTR2C	GDB:378202	5-@HYDROXYTRYPTAMINE RECEPTOR 2C; HTR2C
HYR	GDB:9954625	REGULATOR; HYR
IDS	GDB:120521	MUCOPOLYSACCHARIDOSIS TYPE II
IHG1	GDB:119343	HYPOPLASIA OF, WITH GLAUCOMA; IHG
IL2RG	GDB:134807	INTERLEUKIN-2 RECEPTOR, GAMMA; IL2RG SEVERE COMBINED IMMUNODEFICIENCY DISEASE, X-LINKED, 2; SCIDX2
INDX	GDB:9954657	IMMUNONEUROLOGIC DISORDER, X-LINKED
IP1	GDB:120105	INCONTINENTIA PIGMENTI, TYPE I; IP1
IP2	GDB:120106	INCONTINENTIA PIGMENTI, TYPE II; IP2
JMS	GDB:204055	MENTAL RETARDATION, X-LINKED, WITH GROWTH RETARDATION, DEAFNESS, AND
KAL1	GDB:120116	KALLMANN SYNDROME 1; KAL1
KFSD	GDB:128174	KERATOSIS FOLLICULARIS SPINULOSA DECALVANS CUM OPHIASI; KFSD
L1CAM	GDB:120133	CLASPED THUMB AND MENTAL RETARDATION L1 CELL ADHESION MOLECULE; L1CAM
LAMP2	GDB:125376	LYSOSOME-ASSOCIATED MEMBRANE PROTEIN B; LAMP2; LAMPB
MAA	GDB:119372	MICROPHTHALMIA OR ANOPHTHALMOS, WITH ASSOCIATED ANOMALIES; MAA
MAFD2	GDB:119373	PSYCHOSIS, X-LINKED
MAOA	GDB:120164	MONOAMINE OXIDASE A; MAOA

Gene	GDB Accession ID	OMIM Link
MAOB	GDB:119377	MONOAMINE OXIDASE B; MAOB
5	MCF2	MCF.2 CELL LINE DERIVED TRANSFORMING SEQUENCE; MCF2
	MCS	MENTAL RETARDATION, X-LINKED, SYNDROMIC-4, WITH CONGENITAL CONTRACTURES
	MEAX	X-LINKED, WITH EXCESSIVE AUTOPHAGY; XMEA; MEAX
	MECP2	SYNDROME; RTT
10	MF4	METACARPAL 4-5 FUSION; MF4
	MGC1	MEGALOCORNEA; MGC1; MGCN
	MIC5	SURFACE ANTIGEN, X-LINKED; SAX
	MID1	OPITZ SYNDROME
15	MLLT7	MYELOID/LYMPHOID OR MIXED-LINEAGE LEUKEMIA, TRANSLOCATED TO, 7; MLLT7
	MLS	MICROPHTHALMIA WITH LINEAR SKIN DEFECTS; MLS
	MRSD	MENTAL RETARDATION, SKELETAL DYSPLASIA, AND ABDUCENS PALSY; MRSD
	MRX14	RETARDATION, X-LINKED 14; MRX14
20	MRX1	MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 1; MRX1
	MRX20	MENTAL RETARDATION, X-LINKED 20; MRX20
	MRX2	RETARDATION, X-LINKED NONSPECIFIC, TYPE 2; MRX2
	MRX3	GDP DISSOCIATION INHIBITOR 1; GDI1 MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 3; MRX3
25	MRX40	MENTAL RETARDATION, X-LINKED, WITH HYPOTONIA
	MRXA	MENTAL RETARDATION, X-LINKED NONSPECIFIC, WITH APHASIA; MRXA
30		
35		

Gene	GDB Accession ID	OMIM Link
MSD	GDB:119399	SYNDROME
5	MTM1	MYOTUBULAR MYOPATHY 1; MTM1
	MYCL2	MYCL-RELATED PROCESSED GENE; MYCL2
	MYP1	MYOPIA, X-LINKED; MYP1
	NDP	NORRIE DISEASE; NDP
	NHS	CATARACT-DENTAL SYNDROME
10	NPHL1	NEPHROLITHIASIS, X-LINKED RECESSIVE, WITH RENAL FAILURE; XRN
	NR0B1	ADRENAL HYPOPLASIA, CONGENITAL; AHC
	NSX	SYNDROME; NSX
	NYS1	NYSTAGMUS, X-LINKED; NYS
	NYX	NIGHTBLINDNESS, CONGENITAL STATIONARY, WITH MYOPIA; CSNB1
15	OA1	ALBINISM, OCULAR, TYPE 1; OA1
	OASD	OCULAR, WITH LATE-ONSET SENSORINEURAL DEAFNESS; OASD
	OCRL	LOWE OCULOCEREBRORENAL SYNDROME; OCRL
	ODT1	TEETH, ABSENCE OF
	OFD1	OROFACIODIGITAL SYNDROME 1; OFD1
20	OPA2	OPTIC ATROPHY 2; OPA2
	OPD1	OTOPALATODIGITAL SYNDROME
	OPEM	OPHTHALMOPLEGIA, EXTERNAL, AND MYOPIA; OPEM
	OPN1LW	COLORBLINDNESS, PARTIAL, PROTAN SERIES; CBP
	OPN1MW	COLORBLINDNESS, PARTIAL, DEUTAN SERIES; CBD; DCB
25	OTC	ORNITHINE TRANSCARBAMYLASE DEFICIENCY, HYPERAMMONEMIA DUE TO; OTC
30		
35		

Gene	GDB Accession ID	OMIM Link
P3	GDB:9954667	PROTEIN P3
5 PDHA1	GDB:118895	PYRUVATE DEHYDROGENASE COMPLEX, E1-ALPHA POLYPEPTIDE-1; PDHA1
PDR	GDB:203409	AMYLOIDOSIS, FAMILIAL CUTANEOUS
PFC	GDB:120275	PROPERDIN DEFICIENCY, X-LINKED
10 PFKFB1	GDB:125375	6-@PHOSPHOFRUCTO-2-KINASE; PFKFB1
PGK1	GDB:120282	PHOSPHOGLYCERATE KINASE 1; PGK1
PGK1P1	GDB:120283	PHOSPHOGLYCERATE KINASE 1; PGK1
15 PGS	GDB:128372	DANDY-WALKER MALFORMATION WITH MENTAL RETARDATION, BASAL GANGLIA DISEASE,
PHEX	GDB:120520	HYPOPHOSPHATEMIA, VITAMIN D-RESISTANT RICKETS; HYP
PHKA1	GDB:120285	PHOSPHORYLASE KINASE, ALPHA 1 SUBUNIT (MUSCLE); PHKA1
PHKA2	GDB:127279	GLYCOGEN STORAGE DISEASE VIII
20 PHP	GDB:119494	PANHYPOPITUITARISM; PHP
PIGA	GDB:138138	PHOSPHATIDYLINOSITOL GLYCAN, CLASS A; PIGA
PLP1	GDB:120302	PROTEOLIPID PROTEIN, MYELIN; PLP
POF1	GDB:120716	PREMATURE OVARIAN FAILURE 1; POF1
25 POLA	GDB:120304	POLYMERASE, DNA, ALPHA; POLA
POU3F4	GDB:351386	DEAFNESS, CONDUCTIVE, WITH STAPES FIXATION
PPMX	GDB:9954669	RETARDATION WITH PSYCHOSIS, PYRAMIDAL SIGNS, AND MACROORCHIDISM
30 PRD	GDB:371323	DYSPLASIA, PRIMARY
PRPS1	GDB:120318	PHOSPHORIBOSYL PYROPHOSPHATE SYNTHETASE-I; PRPS1
35 PRPS2	GDB:120320	PHOSPHORIBOSYL PYROPHOSPHATE SYNTHETASE-II; PRPS2

Gene	GDB Accession ID	OMIM Link
PRS	GDB:128368	MENTAL RETARDATION, X-LINKED, SYNDROMIC-2, WITH DYSMORPHISM AND CEREBRAL
5 PRTS	GDB:128367	PARTINGTON X-LINKED MENTAL RETARDATION SYNDROME; PRTS
PSF2	GDB:119519	TRANSPORTER 2, ABC; TAP2
RENBP	GDB:133792	RENIN-BINDING PROTEIN; RENBP
10 RENS1	GDB:9806348	MENTAL RETARDATION, X-LINKED, RENPENNING TYPE
RP2	GDB:120353	RETINITIS PIGMENTOSA-2; RP2
RP6	GDB:125381	PIGMENTOSA-6; RP6
15 RPGR	GDB:118736	RETINITIS PIGMENTOSA-3; RP3
RPS4X	GDB:128115	RIBOSOMAL PROTEIN S4, X-LINKED; RPS4X
RPS6KA3	GDB:365648	RIBOSOMAL PROTEIN S6 KINASE, 90 KD, POLYPEPTIDE 3; RPS6KA3
20 RS1	GDB:119581	RETINOSCHISIS; RS
S11	GDB:120361	ANTIGEN, X-LINKED, SECOND; SAX2
SDYS	GDB:119590	GLYCAN-3; GPC3 SIMPSON DYSMORPHIA SYNDROME; SDYS
25 SEDL	GDB:120372	SPONDYLOEPIPHYSEAL DYSPLASIA, LATE; SEDL
SERPINA7	GDB:120399	THYROXINE-BINDING GLOBULIN OF SERUM; TBG
SH2D1A	GDB:120701	IMMUNODEFICIENCY, X-LINKED PROGRESSIVE COMBINED VARIABLE
30 SHFM2	GDB:226635	SPLIT-HAND/SPLIT-FOOT ANOMALY, X-LINKED
SHOX	GDB:6118451	SHORT STATURE; SS
SLC25A5	GDB:125190	ADENINE NUCLEOTIDE TRANSLOCATOR 2; ANT2
35 SMAX2	GDB:9954643	SPINAL MUSCULAR ATROPHY, X-LINKED LETHAL INFANTILE

Gene	GDB Accession ID	OMIM Link
SRPX	GDB:3811398	RETINITIS PIGMENTOSA-3; RP3
SRS	GDB:136337	MENTAL RETARDATION, X-LINKED, SNYDER-ROBINSON TYPE
STS	GDB:120393	ICHTHYOSIS, X-LINKED
SYN1	GDB:119606	SYNAPSIN I; SYN1
SYP	GDB:125295	SYNAPTOPHYSIN; SYP
10 TAF1	GDB:120573	TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR 2A; TAF2A
TAZ	GDB:120609	CARDIOMYOPATHY, DILATED 3A; CMD3A ENDOCARDIAL FIBROELASTOSIS-2; EFE2
TBX22	GDB:10796448	CLEFT PALATE, X-LINKED; CPX
15 TDD	GDB:119610	MALE PSEUDOHERMAPHRODITISM: DEFICIENCY OF TESTICULAR 17,20-DESMOLASE;
TFE3	GDB:125870	TRANSCRIPTION FACTOR FOR IMMUNOGLOBULIN HEAVY-CHAIN ENHANCER-3; TFE3
20 THAS	GDB:128158	THORACOABDOMINAL SYNDROME; TAS
THC	GDB:125361	THROMBOCYTOPENIA, X-LINKED; THC; XLT
TIMM8A	GDB:119090	DEAFNESS 1, PROGRESSIVE; DFN1
25 TIMP1	GDB:119615	TISSUE INHIBITOR OF METALLOPROTEINASE-1; TIMP1
TKCR	GDB:119616	TORTICOLLIS, KELOIDS, CRYPTORCHIDISM, AND RENAL DYSPLASIA; TKC
TNFSF5	GDB:120632	IMMUNODEFICIENCY WITH INCREASED IgM
30 UBE1	GDB:118954	UBIQUITIN-ACTIVATING ENZYME 1; UBE1
UBE2A	GDB:131647	UBIQUITIN-CONJUGATING ENZYME E2A; UBE2A
WAS	GDB:120736	WISKOTT-ALDRICH SYNDROME; WAS
35 WSN	GDB:125864	PARKINSONISM, EARLY-ONSET, WITH MENTAL RETARDATION

Gene	GDB Accession ID	OMIM Link
WTS	GDB:128373	MENTAL RETARDATION, X-LINKED, SYNDROMIC-6, WITH GYNECOMASTIA AND OBESITY;
5	WWS	WIEACKER SYNDROME
	XIC	X-INACTIVATION-SPECIFIC TRANSCRIPT; XIST
	XIST	X-INACTIVATION-SPECIFIC TRANSCRIPT; XIST
	XK	Xk LOCUS
	XM	XM SYSTEM
10	XS	LUTHERAN SUPPRESSOR, X-LINKED; XS; LUXS
	ZFX	ZINC FINGER PROTEIN, X-LINKED; ZFX
	ZIC3	HETEROTAXY, X-LINKED VISCERAL; HTX1
	ZNF261	MENTAL RETARDATION, X-LINKED; DDX6673E
	ZNF41	ZINC FINGER PROTEIN-41; ZNF41
15	ZNF6	ZINC FINGER PROTEIN-6; ZNF6

Table 25: Genes, Locations and Genetic Disorders on Chromosome Y

Gene	GDB Accession ID	OMIM Link
25	AMELY	AMELOGENIN, Y-CHROMOSOMAL; AMELY
	ASSP6	CITRULLINEMIA
	AZF1	AZOOSPERMIA FACTOR 1; AZF1
	AZF2	AZOOSPERMIA FACTOR 2; AZF2
	DAZ	DELETED IN AZOOSPERMIA; DAZ
30	GCY	CONTROL, Y-CHROMOSOME INFLUENCED; GCY
	RPS4Y	RIBOSOMAL PROTEIN S4, Y-LINKED; RPS4Y
	SMCY	HISTOCOMPATIBILITY Y ANTIGEN; HY; HYA
	SRY	SEX-DETERMINING REGION Y; SRY

Gene	GDB Accession ID	OMIM Link
ZFY	GDB:120503	ZINC FINGER PROTEIN, Y-LINKED; ZFY

5 Table 26: Genes, Locations and Genetic Disorders in Unknown or Multiple Locations

Gene	GDB Accession ID	OMIM Link
ABAT	GDB:581658	GAMMA-AMINOBUTYRATE TRANSAMINASE
AEZ	GDB:128360	ACRODERMATITIS ENTEROPATHICA, ZINC-DEFICIENCY TYPE; AEZ
AFA	GDB:265277	FILIFORME ADNATUM AND CLEFT PALATE
AFD1	GDB:265292	DYSOSTOSIS, TREACHER COLLINS TYPE, WITH LIMB ANOMALIES
AGS1	GDB:10795417	ENCEPHALOPATHY, FAMILIAL INFANTILE, WITH CALCIFICATION OF BASAL GANGLIA
ASAHI	GDB:6837715	FARBER LIPOGRANULOMATOSIS
ASD1	GDB:6276019	ATRIAL SEPTAL DEFECT; ASD
ASMT	GDB:136259	CETYLSEROTONIN METHYLTRANSFERASE; ASMT ACETYL SEROTONIN METHYLTRANSFERASE, Y-CHROMOSOMAL; ASMTY; HIOMTY
BCH	GDB:118758	CHOREA, HEREDITARY BENIGN; BCH
CCAT	GDB:118738	CATARACT, CONGENITAL OR JUVENILE
CECR9	GDB:10796163	CAT EYE SYNDROME; CES
CEPA	GDB:581848	CONTROL, CONGENITAL FAILURE OF
CHED2	GDB:9957389	CORNEAL DYSTROPHY, CONGENITAL HEREDITARY
CLA1	GDB:119781	CEREBELLOPARENCHYMAL DISORDER III
CLA3	GDB:128453	CEREBELLOPARENCHYMAL DISORDER I; CPD I
CLN4	GDB:125229	CEROID-LIPOFUSCINOSIS, NEURONAL 4; CLN4
CPO	GDB:119070	COPROPORPHYRIA

Gene	GDB Accession ID	OMIM Link
5 CSF2RA	GDB:118777	COLONY STIMULATING FACTOR 2 RECEPTOR, ALPHA; CSF2RA GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR, ALPHA SUBUNIT,
10 CTS1	GDB:118779	CARPAL TUNNEL SYNDROME; CTS; CTS1
DIH1	GDB:132645	FACTOR D
DWS	GDB:439243	DIAPHRAGMATIC
DWS	GDB:128371	SYNDROME; DWS
DYT2	GDB:118788	DYSTONIA MUSCULORUM DEFORMANS 2; DYT2
DYT4	GDB:433751	DYSTONIA MUSCULORUM DEFORMANS 4; DYT4
15 EBR3	GDB:118739	EPIDERMOLYSIS BULLOSA DYSTROPHICA NEUROTROPHICA
ECT	GDB:128640	CENTRALOPATHIC EPILEPSY
EEF1A1L14	GDB:1327185	PROSTATIC CARCINOMA ONCOGENE PTI-1
20 EYCL2	GDB:4642815	EYE COLOR-3; EYCL3
FA1	GDB:118795	FANCONI ANEMIA, COMPLEMENTATION GROUP A; FACA
FANCB	GDB:9864269	FANCONI PANCYTOPENIA, TYPE 2
25 GCSH	GDB:126842	HYPERGLYCINEMIA, ISOLATED NONKETOTIC, TYPE III; NKH3
GCSL	GDB:132139	ISOLATED NONKETOTIC, TYPE IV; NKH4
GDF5	GDB:433948	CARTILAGE-DERIVED MORPHOGENETIC PROTEIN 1
30 GIP	GDB:119985	GASTRIC INHIBITORY POLYPEPTIDE; GIP
GTS	GDB:118807	GILLES DE LA TOURETTE SYNDROME; GTS
HHG	GDB:118740	HYPERGONADOTROPIC HYPOGONADISM; HHG
HMI	GDB:265275	OF ITO; HMI
35 HOAC	GDB:118812	DEAFNESS, CONGENITAL, AUTOSOMAL RECESSIVE

	Gene	GDB Accession ID	OMIM Link
	HOKPP2	GDB:595535	HYPOKALEMIC PERIODIC PARALYSIS, TYPE II; HOKPP2
5	HRPT1	GDB:125252	HYPERPARATHYROIDISM, FAMILIAL PRIMARY
	HSD3B3	GDB:676973	GIANT CELL HEPATITIS, NEONATAL
	HTC1	GDB:265286	HYPERTRICHOSIS UNIVERSALIS CONGENITA, AMBRAS TYPE; HTC1
10	HV1S	GDB:9955009	HERPES VIRUS SENSITIVITY; HV1S
	ICR1	GDB:127785	LAMELLAR, AUTOSOMAL DOMINANT FORM
	ICR5	GDB:127789	ICHTHYOSIS CONGENITA, HARLEQUIN FETUS TYPE
15	IL3RA	GDB:128985	INTERLEUKIN-3 RECEPTOR, ALPHA; IL3RA INTERLEUKIN-3 RECEPTOR, Y-CHROMOSOMAL; IL3RA
	KAL2	GDB:265288	KALLMANN SYNDROME 2; KAL2
	KMS	GDB:118827	SYNDROME; KMS
	KRT18	GDB:120127	KERATIN 18; KRT18
20	KSS	GDB:9957718	KEARNS-SAYRE SYNDROME; KSS
	LCAT	GDB:119359	FISH-EYE DISEASE; FED LECITHIN:CHOLESTEROL ACYLTRANSFERASE DEFICIENCY
	LIMM	GDB:9958161	MYOPATHY, MITOCHONDRIAL, LETHAL INFANTILE; LIMM
25	MANBB	GDB:125262	MANNOSIDOSIS, BETA; MANB1
	MCPH2	GDB:9863035	MICROCEPHALY; MCT
	MEB	GDB:599557	DISEASE
	MELAS	GDB:9955855	MELAS SYNDROME
30	MIC2	GDB:120184	SURFACE ANTIGEN MIC2; MIC2; CD99 MIC2 SURFACE ANTIGEN, Y-CHROMOSOMAL; MIC2Y
	MPFD	GDB:439372	CONGENITAL, WITH FIBER-TYPE DISPROPORTION
	MS	GDB:229116	SCLEROSIS; MS
35	MSS	GDB:118743	MARINESCO-SJOGREN SYNDROME; MSS

Gene	GDB Accession ID	OMIM Link
MTATP6	GDB:118897	ATP SYNTHASE 6; MTATP6
5 MTCO1	GDB:118900	COMPLEX IV, CYTOCHROME c OXIDASE SUBUNIT I; MTCO1; COI
MTCO3	GDB:118902	CYTOCHROME c OXIDASE III; MTCO3
MTCYB	GDB:118906	COMPLEX III, CYTOCHROME b SUBUNIT
10 MTND1	GDB:118911	COMPLEX I, SUBUNIT ND1; MTND1
MTND2	GDB:118912	COMPLEX I, SUBUNIT ND2; MTND2
MTND4	GDB:118914	COMPLEX I, SUBUNIT ND4; MTND4
MTND5	GDB:118916	COMPLEX I, SUBUNIT ND5; MTND5
MTND6	GDB:118917	COMPLEX I, SUBUNIT ND6; MTND6
15 MTRNR1	GDB:118920	RIBOSOMAL RNA, MITOCHONDRIAL, 12S; MTRNR1
MTRNR2	GDB:118921	RIBOSOMAL RNA, MITOCHONDRIAL, 16S; MTRNR2
MTTE	GDB:118926	TRANSFER RNA, MITOCHONDRIAL, GLUTAMIC ACID; MTTE
20 MTTG	GDB:118933	TRANSFER RNA, MITOCHONDRIAL, GLYCINE; MTTG
MTTI	GDB:118935	TRANSFER RNA, MITOCHONDRIAL, ISOLEUCINE; MTTI
25 MTTK	GDB:118936	MERRF SYNDROME TRANSFER RNA, MITOCHONDRIAL, LYSINE; MTTK
MTTL1	GDB:118937	MERRF SYNDROME TRANSFER RNA, MITOCHONDRIAL, LEUCINE, 1; MTTL1
MTTL2	GDB:118938	TRANSFER RNA, MITOCHONDRIAL, LEUCINE, 2; MTTL2
30 MTTN	GDB:118940	TRANSFER RNA, MITOCHONDRIAL, ASPARAGINE; MTTN
MTTP	GDB:118941	TRANSFER RNA, MITOCHONDRIAL, PROLINE; MTTP
MTTS1	GDB:118944	TRANSFER RNA, MITOCHONDRIAL, SERINE, 1; MTTS1

Gene	GDB Accession ID	OMIM Link
NAMSD	GDB:681237	NEUROPATHY, MOTOR-SENSORY, TYPE II, WITH DEAFNESS AND MENTAL RETARDATION
5	NODAL	NODAL, MOUSE, HOMOLOG OF
	OCD1	DISORDER-1; OCD1
	OPD2	SYNDROME
10	PCK2	PHOSPHOENOLPYRUVATE CARBOXYKINASE 2, MITOCHONDRIAL; PCK2
	PCLD	POLYCYSTIC LIVER DISEASE; PLD
	PCOS1	STEIN-LEVENTHAL SYNDROME
15	PFKM	GLYCOGEN STORAGE DISEASE VII
	PKD3	KIDNEY DISEASE 3, AUTOSOMAL DOMINANT; PKD3
	PRCA1	PROSTATE CANCER; PRCA1
20	PRO1	
	PROP1	PROPHET OF PIT1, MOUSE, HOMOLOG OF; PROP1
	RBS	ROBERTS SYNDROME; RBS
25	RFXAP	REGULATORY FACTOR X-ASSOCIATED PROTEIN; RFXAP
	RP	RETINITIS PIGMENTOSA-8
	SLC25A6	ADENINE NUCLEOTIDE TRANSLOCATOR 3; ANT3 ADENINE NUCLEOTIDE TRANSLOCATOR 3, Y-CHROMOSOMAL; ANT3Y
30	SPG5B	SPASTIC PARAPLEGIA-5B, AUTOSOMAL RECESSIVE; SPG5B
	STO	CEREBRAL GIGANTISM
	SUOX	SULFOCYSTEINURIA
35	TC21	ONCOGENE TC21
	THM	FAMILIAL
	TST	RHODANESE; RDS
35	TTD	TRICHOThIODYSTROPHY; TTD

The present invention is not to be limited in scope by the specific  
embodiments described herein. Indeed, various modifications of the invention in addition  
to those described will become apparent to those skilled in the art from the foregoing  
description and accompanying figures. Such modifications are intended to fall within the  
5 scope of the appended claims.

Various publications are cited herein, the disclosures of which are  
incorporated by reference in their entireties.

10

15

20

25

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

1. A method for identifying a test compound that modulates premature translation termination or nonsense-mediated mRNA decay, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed; (b) separating the detectably labeled target RNA:support-attached test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds, and (c) determining a structure of the test compound of the RNA:support-attached test compound complex.
2. The method of paragraph 1 in which the target RNA molecule contains regions of 28S rRNA or analogs thereof.
3. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
4. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries including, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.
5. The method of paragraph 1 in which screening a library of test compounds preferably comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions

6. The method of paragraph 5 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNA.

5

7. The method of paragraph 5 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 10 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

10

8. Any method that detects an altered physical property of a target 15 nucleic acid complexed to a test compound attached to a solid support from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. Methods such as flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave are used for the separation of the complexed and non-complexed target nucleic acids. 20

25

9. The structure of the substantially one type of test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds. In an embodiment wherein the combinatorial libraries are peptide or peptide-based libraries, Edman degradation is used to determine the structure of the test compounds.

30

35

## WHAT IS CLAIMED IS:

1. A method for identifying a test compound that modulates premature translation termination or nonsense-mediated mRNA decay, comprising the steps of:

5 (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed;

10 (b) separating the detectably labeled target RNA:support-attached test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds by flow cytometry; and

15 (c) determining a structure of the substantially one type of test compound of the RNA:support-attached test compound complex by mass spectroscopy.

20

25

30

35

ABSTRACT

The present invention relates to a method for screening and identifying test compounds that modulate premature translation termination and/or nonsense-mediated messenger ribonucleic acid ("mRNA") by interacting with a preselected target ribonucleic acid ("RNA"). In particular, the present invention relates to identifying test compounds that bind to regions of the 28S ribosomal RNA ("rRNA") and analogs thereof. Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

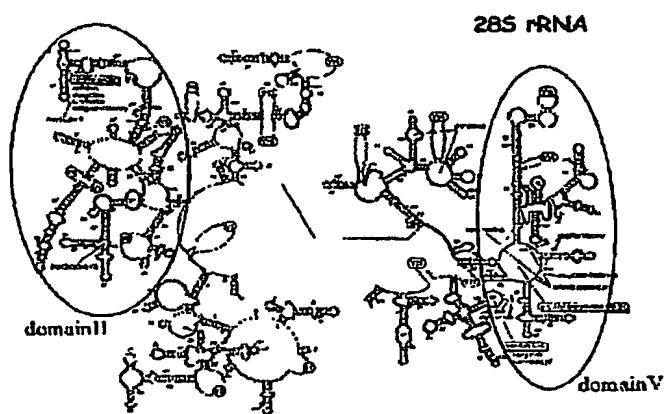
20

25

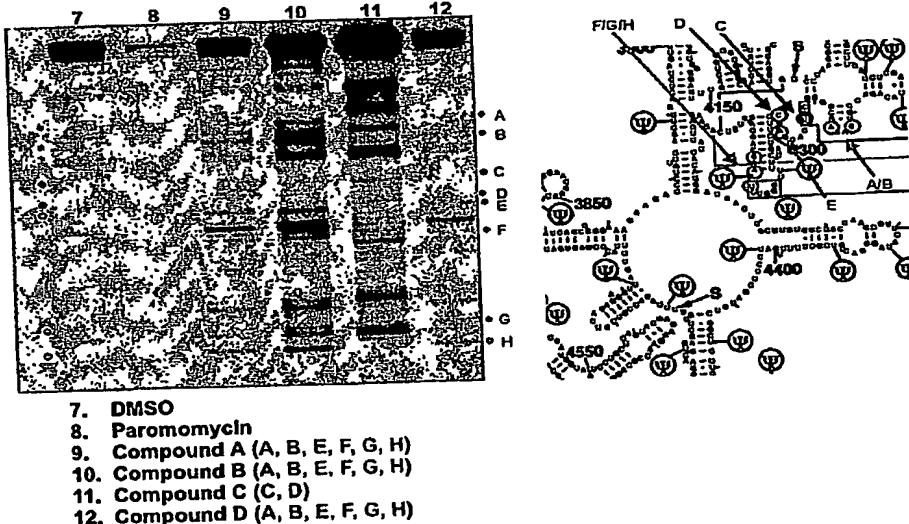
30

35

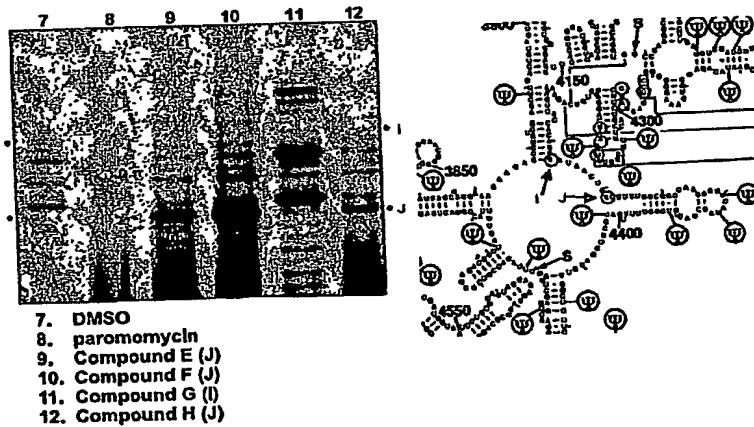
**Figure 1**  
**Sheet 1/6**  
**Attorney Docket No. 10589-023**



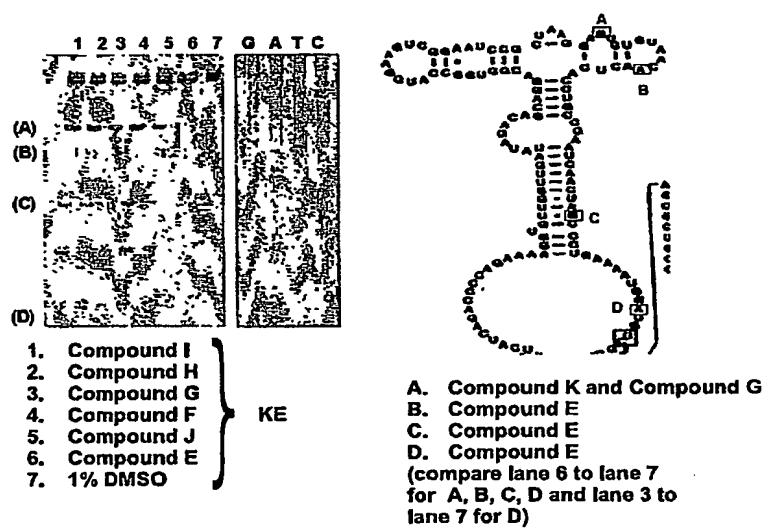
**Figure 2**  
**Sheet 2/6**



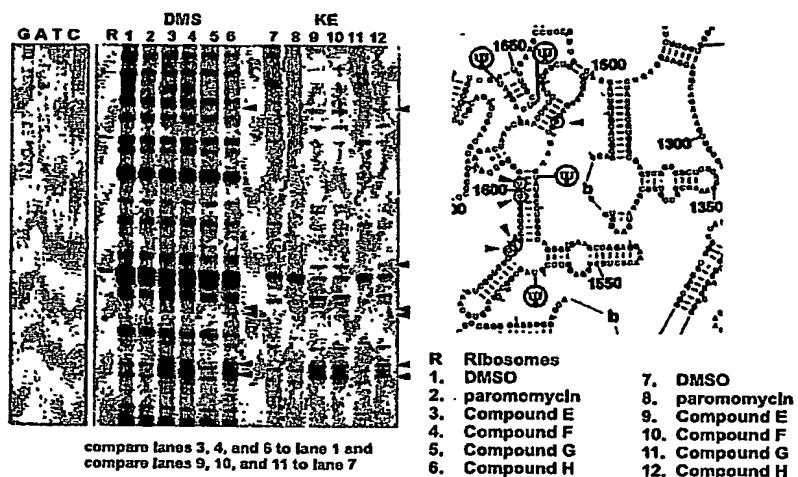
NY2-1320481.1

**Figure 3****Sheet 3/6****Attorney Docket No. 10589-023**

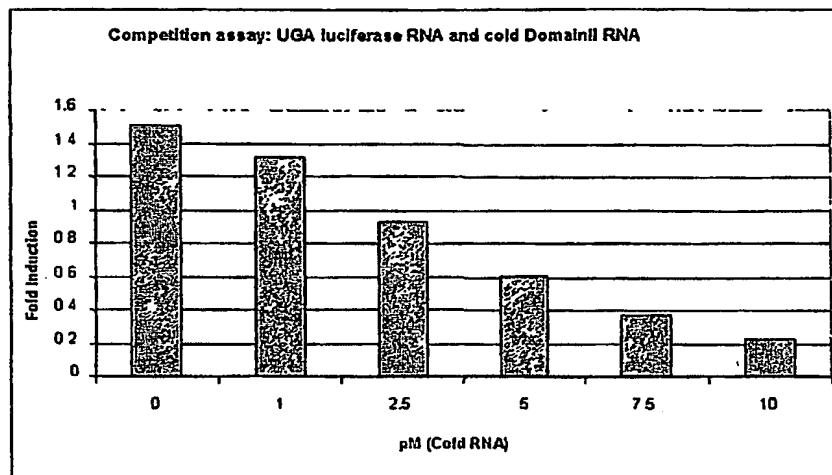
**Figure 4**  
**Sheet 4/6**  
**Attorney Docket No. 10589-023**



**Figure 5**  
**Sheet 5/6**  
**Attorney Docket No. 10589-023**



**Figure 6**  
**Sheet 6/6**  
**Attorney Docket No. 10589-023**



PI 1057485

REC'D 05 SEP 2003

WIPO PCT

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office

August 29, 2003

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM  
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK  
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT  
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A  
FILING DATE.

APPLICATION NUMBER: 60/398,344

FILING DATE: July 24, 2002

RELATED PCT APPLICATION NUMBER: PCT/US03/23075

By Authority of the  
COMMISSIONER OF PATENTS AND TRADEMARKS

*L. Edele*

L. EDELEN  
Certifying Officer

PRIORITY  
DOCUMENT

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

07-25-02 60398344-UV-24-US A  
Express Mail No.: EV 058 489 450 US

07/24/02  
05879 U.S. PRO  
60/398344  
07/24/02

## COVER SHEET FOR PROVISIONAL APPLICATION FOR PATENT

United States Commissioner for Patents  
Box PROVISIONAL PATENT APPLICATION  
Washington, DC 20231

Sir:

This is a request for filing a PROVISIONAL APPLICATION under 37 CFR 1.53(c).

		Docket Number	10589-021-888	Type a plus sign (+) inside this box -	+
INVENTOR(S) APPLICANT(S)					
LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)		
TITLE OF THE INVENTION (280 characters max)					
METHODS FOR IDENTIFYING SMALL MOLECULES THAT MODULATE PREMATURE TRANSLATION TERMINATION AND NONSENSE MEDIATED mRNA DECAY					
PENNIE & EDMONDS LLP CORRESPONDENCE ADDRESS 20583					
ENCLOSED APPLICATION PARTS (check all that apply)					
<input checked="" type="checkbox"/> Specification	Number of Pages	154	<input type="checkbox"/> Applicant claims small entity status, see 37 CFR § 1.27		
<input checked="" type="checkbox"/> Drawing(s)	Number of Sheets	11	<input type="checkbox"/> Other (specify)		
METHOD OF PAYMENT (check one)					
<input type="checkbox"/> A check or money order is enclosed to cover the Provisional filing fees.				ESTIMATED PROVISIONAL FILING FEE AMOUNT	<input checked="" type="checkbox"/> \$160 <input type="checkbox"/> \$80
<input checked="" type="checkbox"/> The Commissioner is hereby authorized to charge the required filing fee to Deposit Account Number 16-1150					

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government  
 No.  Yes, the name of the U.S. Government agency and the Government contract number are \_\_\_\_\_

Respectfully submitted,

Signature Laura A. Coruzzi REGISTRATION NO. 30,742 Date July 24, 2002  
Laura A. Coruzzi  
PENNIE & EDMONDS LLP

by: Deborah L. den Reg No. 50,940

Additional inventors are being named on separately numbered sheets attached hereto

Total number of cover sheet pages

166

**PROVISIONAL APPLICATION FILING ONLY**

EXPRESS MAIL NO. EV 058 489 450 US  
ATTORNEY DOCKET NO. 10589-021-888

**METHODS FOR IDENTIFYING SMALL MOLECULES THAT  
MODULATE PREMATURE TRANSLATION TERMINATION  
AND NONSENSE MEDIATED mRNA DECAY**

5

**1. INTRODUCTION**

The present invention relates to a method for screening and identifying test compounds that modulate premature translation termination and/or nonsense-mediated messenger ribonucleic acid ("mRNA") decay by interacting with a preselected target ribonucleic acid ("RNA"). In particular, the present invention relates to identifying test compounds that bind to regions of the 28S ribosomal RNA ("rRNA") and analogs thereof. Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

20

**2. BACKGROUND OF THE INVENTION**

Protein synthesis encompasses the processes of translation initiation, elongation, and termination, each of which has evolved to occur with great accuracy and has the capacity to be a regulated step in the pathway of gene expression. Recent studies, including those suggesting that events at termination may regulate the ability of ribosomes to recycle to the start site of the same mRNA, have underscored the potential of termination to regulate other aspects of translation. The RNA triplets UAA, UAG, and UGA are non-coding and promote translational termination. Termination starts when one of the three termination codons enters the A site of the ribosome signaling the polypeptide chain release factors to bind and recognize the termination signal. Subsequently, the ester bond between the 3' nucleotide of the transfer RNA ("tRNA") located in the ribosome's P site and the nascent polypeptide chain is hydrolyzed, the completed polypeptide chain is released, and the ribosome subunits are recycled for another round of translation.

Nonsense-mediated mRNA decay is a surveillance mechanism that minimizes the translation and regulates the RNA stability of nonsense RNAs that contain chain termination mutations (see, e.g., Hentze & Kulozik, 1999, Cell 96:307-310; Culbertson, 1999, Trends in Genetics 15:74-80; Li & Wilkinson, 1998, Immunity 8:135-

141; and Ruiz-Echevarria *et al.*, 1996, Trends in Biological Sciences, 21:433-438). Chain 5 termination mutations are caused when a base substitution or frameshift mutation changes a codon into a termination codon, *i.e.*, a premature stop codon that causes translational termination. In nonsense-mediated mRNA decay, mRNAs with premature stop codons are frequently subject to degradation. A truncated protein is produced as a result of the translation apparatus prematurely terminating at the stop codon.

Certain classes of known antibiotics have been characterized and found to interact with RNA. For example, the antibiotic thiostrepton binds tightly to a 60-mer from 10 ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, *Ann. Rev. Microbiol.* 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron 15 splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, *Nature (London)* 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, *RNA* 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, *EMBO J.* 10:3099-3103). An 20 oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, *Nature* 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in 25 mitochondrial rRNA (Hutchin *et al.*, 1993, *Nucleic Acids Res.* 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and 30 tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Aminoglycosides have also been found to promote nonsense suppression (see, *e.g.*, Bedwell *et al.*, 1997, *Nat. Med.* 3:1280-1284 and Howard *et al.*, 1996, *Nat. Med.* 2:467-469). Nonsense mutations cause approximately 10 to 30 percent of the individual cases of virtually all inherited diseases. Although nonsense mutations inhibit the synthesis 35 of a full-length protein to one percent or less of wild-type levels, minimally boosting the expression levels of the full-length protein to between five and fifteen percent of normal

levels can eliminate or greatly reduce the severity of the disease. Clinical approaches that target the translation termination event to promote nonsense suppression have recently been described for model systems of cystic fibrosis and muscular dystrophy. Gentamicin is an 5 aminoglycoside antibiotic that causes translational misreading and allowed the insertion of an amino acid at the site of the nonsense codon in models of cystic fibrosis, Hurlers Syndrome, and muscular dystrophy (see, e.g., Barton-Davis *et al.*, 1999, *J. Clin. Invest.* 104:375-381). These results strongly suggest that drugs that promote nonsense suppression by altering translation termination efficiency of a premature termination codon can be 10 therapeutically valuable in the treatment of diseases caused by nonsense mutations.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

### **3. SUMMARY OF THE INVENTION**

15 The present invention relates to methods for identifying compounds that modulate translation termination and/or nonsense-mediated mRNA decay by identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. In particular, the present invention relates to identifying test compounds that bind 20 to regions of the 28S rRNA and analogs thereof. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements (*i.e.*, regions of the 28S rRNA and analogs thereof) are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used 25 to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled 30 tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the 35 electrophoretic, chromatographic, or thermostable properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy,

surface plasmon resonance, mass spectrometry, scintillation proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to 28S rRNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high-throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to 28S rRNA preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

### 3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not

limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, 28S rRNA and structural analogs thereof, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1

5 As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each 10 occupy a unique three-dimensional space.

10 As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ ), dyes, fluorescent dyes, electron-dense reagents, enzymes 15 and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of 20 the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, 25 biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der 30 Waals or hydrogen bonds to the molecule to be detected.

35 As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and

about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is 5 detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or 10 phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in 15 nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, 20 phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino 25 moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

30 By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

35

#### 4. DESCRIPTION OF DRAWINGS

FIG. 1. The human 28S rRNA. Domains II and V are circled.

**FIG. 2.** Gel retardation analysis to detect peptide-RNA interactions. In 20  $\mu$ l reactions containing increasing concentrations of Tat<sub>47-58</sub> peptide (0.1  $\mu$ M, 0.2  $\mu$ M, 0.4  $\mu$ M, 0.8  $\mu$ M, 1.6  $\mu$ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. 10 ml of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.

FIG. 3. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20  $\mu$ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1  $\mu$ g/ml, 10  $\mu$ g/ml, 50  $\mu$ g/ml, 500  $\mu$ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90°C for 2 min and allowed to cool slowly to 24°C. Then 10  $\mu$ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4°C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane.

FIG. 4. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20  $\mu$ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 3.

FIG. 5. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of  $MgCl_2$ . Reaction mixtures containing gentamicin (1 mg/ml, 100  $\mu$ g/ml, 10  $\mu$ g/ml, 1  $\mu$ g/ml, 0.1  $\mu$ g/ml, and 10 ng/ml) were treated as described in Figure 3 except that the TKM buffer does not contain  $MgCl_2$ .

FIG. 6. Gel retardation analysis to detect peptide-RNA interactions. In reactions containing increasing concentrations of Tat<sub>47-58</sub> peptide (0.1  $\mu$ M, 0.2  $\mu$ M, 0.4  $\mu$ M, 0.8  $\mu$ M, 1.6  $\mu$ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. The reactions were loaded onto a SCE9610

automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

5 FIG. 7.

10 FIG. 8.

15 FIG. 9.

20 FIG. 10.

25 FIG. 11.

Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain V of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain V of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

Small molecules involved in nonsense suppression alter the chemical footprinting pattern in Domain II (GTPase Center) of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with the chemical modifying agents kethoxal (KE) and dimethyl sulfate (DMS, not shown). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

Small molecules involved in nonsense suppression alter the chemical footprinting pattern of domain II of the 28S rRNA. 100 pmol of ribosomes were incubated with 100  $\mu$ M compound, followed by treatment with chemical modifying agents dimethyl sulfate (DMS) and kethoxal (KE). Following chemical modification, rRNA was prepared and analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to rRNA. A sequencing reaction was run in parallel as a marker.

A specific region of Domain II can compete for compound binding and prevents nonsense suppression in vitro. The in vitro nonsense suppression

5

assay was performed using a luciferase construct with a UGA nonsense mutation. 0.1  $\mu$ M compound was present in the reaction to induce nonsense suppression. Competitor RNA corresponding to Domain II was added at the indicated concentrations (0, 1, 2.5, 5, 7.5, 10 pM) to titrate the small molecule and prevent nonsense suppression.

### 5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. In particular, the present invention relates to identifying test compounds that bind to regions of the 28S rRNA and analogs thereof. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used. The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to 28S rRNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of 28S rRNA.

Section 5.1 describes examples of 28S rRNA and analogs thereof that can be used as preselected target RNAs. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

### **5.1. 28S rRNA and Analogs Thereof**

The ribosome is a 2.5-MDa ribonucleoprotein complex involved in the decoding of genetic material from mRNA to proteins. A combination of biophysical and biochemical analysis have provided three dimensional models of the ribosome as well as

detailed analyses into the mechanism of the individual steps in translation (see, *e.g.*, Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716; Cate *et al.*, 1999, *Science* 285(5436):2095-2104; and Ban *et al.*, 2000, *Science*.289(5481): 905-920.).

5 The 28S rRNA is one of the ribosomal RNA components of the 60S subunit of eucaryotic ribosomes. The 28S rRNA sequences are conserved when expressed as mature rRNAs, although the 28S rRNA contains variable sequence tracts that are interspersed among conserved core sequences and lacking in the counterpart bacterial 23S rRNA (see, *e.g.*, Hancock & Dover, 1988, *Mol. Biol. Evol.* 5:377-391). A diagram of the 10 28S rRNA is presented in Figure 1, with domains II and V circled. As indicated in Figure 1, a GTPase center has been mapped to domain II and the peptidyl transferase center has been mapped to domain V.

15 Compounds that interact in these regions or modulate local changes within these domains of the ribosome (*e.g.*, alter base pairing interactions, base modification or modulate binding of trans-acting factors that bind to these regions) have the potential to modulate translation termination. These regions, *i.e.*, domains II and V are conserved from prokaryotes to eukaryotes, but the role of these regions in modulating translation 20 termination has not been realized in eukaryotes. In bacteria, when a short RNA fragment, complementary to the *E. coli* 23S rRNA segment comprising nucleotides 735 to 766 (in domain II), is expressed *in vivo*, suppression of UGA nonsense mutations, but not UAA or UAG, results (Chernyaeva *et al.*, 1999, *J Bacteriol* 181:5257-5262). Other regions of the 23S rRNA in *E. coli* have been implicated in nonsense suppression including the GTPase center in domain II (nt 1034-1120; Jemiolo *et al*, 1995, *Proc. Nat. Acad. Sci.* 92:12309-12313).

25 Genetic mutations in bacteria have also identified rRNA mutations that either increase the level of frameshifting in the *trpE* or the suppression of a nonsense mutations in the *trpA* gene (reviewed in Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716). The frameshifting mutations mapped to domains IV and V of the 23S rRNA. Disruption of the interaction of the CCA end of the tRNA with the peptidyl transferase 30 center of the ribosome has been demonstrated to result in an increased translational error frequency (reviewed in Green & Noller, 1997, *Annu. Rev. Biochem.* 66:679-716).

35 Regions of the 28S rRNA involved in frameshifting, nonsense mutation suppression, GTPase activity, or peptidyl transferase are attractive target RNAs to identify compounds that modulate premature translation termination and/or nonsense mediated decay. The interference of a test compound with one or more of these functions could potentially mediate translation termination by interfering with premature translation

termination. Without being bound by theory, a test compound could potentially mediate translation termination by causing readthrough of a premature translation codon, therefore allowing the synthesis of the full-length protein.

5 In a preferred embodiment, the target RNA comprises a region of 28S rRNA corresponding to domain II (see, e.g., nucleotides 1310 to 2333 of accession number M11167) or domain V of 28S rRNA (see, e.g., nucleotides 3859 to 4425 of accession number M11167) or an analog thereof. It will become apparent to one of skill in the art that an analog of the 28S rRNA has an analogous structure and function to native 28S rRNA.

10 For example, an analog of human 28S rRNA includes, but is not limited to, a human 28S rRNA retropseudogene (see, e.g., Wang *et al.*, 1997, Gene 196:105-111, Accession Number L20636). Regions corresponding to domain II or domain V of the 28S rRNA pseudogene could be used as target RNAs in the present invention. In a preferred embodiment, the 28S rRNA is a human 28S rRNA, although the teachings of the present invention are applicable to mammals.

15 Synthesis of the target RNAs, *i.e.*, regions of 28S rRNA, can be performed by methods known to one of skill in the art (see, e.g., Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York and Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II). In a preferred embodiment, the target RNAs are cloned as DNAs downstream of a promoter, such as but not limited to T7, T3, or Sp6 promoters, and *in vitro* transcribed with the corresponding polymerase. A detectable label can be incorporated into the *in vitro* transcribed RNA or alternatively, the target RNA is end-labeled (see Section 5.2 *infra*). Alternatively, the target RNA can be amplified by polymerase chain reaction with a primer containing an RNA promoter and subsequently *in vitro* transcribed, as described in U.S. Patent No. 6,271,002, which is incorporated by reference in its entirety.

### 5.2. Detectably Labeled Target RNAs

30 Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (e.g., fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (e.g., Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, etc.), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (e.g., <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C, <sup>32</sup>P, <sup>33</sup>P, etc.), enzymes (e.g., horse radish peroxidase, alkaline phosphatase etc.), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm.

Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (e.g., oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (e.g., the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (see, e.g., Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, *Genetic Engineering* 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, *Biochemistry*, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, e.g., trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated

intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate. Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a

5 mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

10 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an 15 aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using 20 phosphatase treatment followed by end-labeling with ATP-<sup>32</sup>S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

25 A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

30 After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After 35 purification, the target RNA is refolded into its native conformation, preferably by heating

to approximately 85-95°C and slowly cooling to room temperature in a buffer, e.g., a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag<sub>2</sub>S, ZnS, CdS, CdTe, Au, or TiO<sub>2</sub>. Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

20

### 5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphoric acids and  $\alpha$ -amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and

5 diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

10 Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited ), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

15 In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, *J. Comb. Chem.* 3:477-517; Hall *et al.*, 2001, *J. Comb. Chem.* 3:125-150; Dolle, 2000, *J. Comb. Chem.* 2:383-433; Dolle, 1999, *J. Comb. Chem.* 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt , eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

30 Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No.

6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

5 In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit 10 reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may 15 mimic the effects of non-peptides or peptides. In contrast to solid-phase synthesis of combinatorial compound libraries, liquid-phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid-phase 20 synthesis (Egner *et al.*, 1995, *J.Org. Chem.* 60:2652; Anderson *et al.*, 1995, *J. Org. Chem.* 60:2650; Fitch *et al.*, 1994, *J. Org. Chem.* 59:7955; Look *et al.*, 1994, *J. Org. Chem.* 49:7588; Metzger *et al.*, 1993, *Angew. Chem., Int. Ed. Engl.* 32:894; Youngquist *et al.*, 1994, *Rapid Commun. Mass Spect.* 8:77; Chu *et al.*, 1995, *J. Am. Chem. Soc.* 117:5419; 25 Brummel *et al.*, 1994, *Science* 264:399; Stevanovic *et al.*, 1993, *Bioorg. Med. Chem. Lett.* 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis 25 method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, *Chem. Rev.* 97:41-448; Ohlmeyer *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of 30 test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, *Chem. Rev.* 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

35 As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid

support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY),

5 polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g.,

10 TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, i.e., it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestable.

15 In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

20 In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

25 Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, 30 when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind

most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in its entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

#### 5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The

aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1

5 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. Without being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl<sub>2</sub> stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

10 The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-15 laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl  $\alpha$ -D-20 glucopyranoside, n-decyl  $\beta$ -D-maltopyranoside, n-dodecyl  $\beta$ -D-maltoside, n-octyl  $\beta$ -D-glucopyranoside, sorbitan esters, n-tetradecyl  $\beta$ -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxy polyethoxyethanol (NP-40), and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

30 Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, e.g., bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from 35 about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably

performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible or fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

30

### 5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon

resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation.

5

### 5.5.1. Electrophoresis

Methods for separation of the complex of a target RNA bound to a test compound from the unbound RNA comprises any method of electrophoretic separation, including but not limited to, denaturing and non-denaturing polyacrylamide gel electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, agarose gel electrophoresis, and capillary electrophoresis.

10 In a preferred embodiment, an automated electrophoretic system comprising a capillary cartridge having a plurality of capillary tubes is used for high-throughput screening of test compounds bound to target RNA. Such an apparatus for performing automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430; 15 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by reference in their entireties.

20 The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by reference in its entirety, allows one to simultaneously introduce samples into a plurality of capillary tubes directly from microtiter trays having a standard size. U.S. Patent No. 5,885,430 discloses a disposable capillary cartridge which can be cleaned between 25 electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each capillary tube is retained in a mounting plate, the first ends collectively forming an array in the mounting plate. The spacing between the first ends corresponds to the spacing between the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of the capillary tubes can simultaneously be dipped into the samples present in the tray's wells. The cartridge is provided with a second mounting plate in which the second ends of the 30 capillary tubes are retained. The second ends of the capillary tubes are arranged in an array which corresponds to the wells in the microtiter tray, which allows for each capillary tube to be isolated from its neighbors and therefore free from cross-contamination, as each end is dipped into an individual well.

35 Plate holes may be provided in each mounting plate and the capillary tubes inserted through these plate holes. In such a case, the plate holes are sealed airtight so that the side of the mounting plate having the exposed capillary ends can be pressurized. Application of a positive pressure in the vicinity of the capillary openings in this mounting

plate allows for the introduction of air and fluids during electrophoretic operations and also can be used to force out gel and other materials from the capillary tubes during reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The system also has an unit connected to analyze the modulation frequencies of light intensity

received by the photoreceptor so that velocities of the bands along the channel are determined, which allows the materials to be analyzed.

The system disclosed in U.S. Patent No. 5,699,157 also provides for a method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA: test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U<sub>2</sub>, CL<sub>3</sub>, T<sub>1</sub>, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate, [2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, Biochem. 37:2207-2214; Latham & Cech, 1989, Science 245:276-282; and Sambrook *et al.*, 2001, in Molecular Cloning: A

5 Laboratory Manual, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties. The specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore determined by the three-dimensional structure of the RNA.

10 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify and characterize the binding of small molecules to RNA as described, for example, by Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

15 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A 20 change in the cleavage pattern directly indicates that the test compound binds to the target nucleic acid. Multiple test compounds can be examined both in parallel and serially.

25 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

### 5.5.2. Fluorescence Spectroscopy

30 In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (e.g., the labeled target nucleic acid of the present invention), can be used to read reaction results without electrophoretic separation of the samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 35 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of

which are incorporated by reference in their entireties. The application of fluorescence polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

5 In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence 10 quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, *Anal. Biochem.* 261(2):183-90).

15 In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a 20 binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench 25 fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test 30 compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, *Bioorg. Med. Chem. Lett.* 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein fluorophore and a tetramethylrhodamine on the other end. The conformational change of 35 the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

35 In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives,

bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

5 If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, 10 available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

15 In yet another embodiment, homogeneous time-resolved fluorescence ("HTRF") techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, *J. Biomol. Screening* 4:303-307; Mathis, 1999, *J. Biomol. Screening* 4:309-313). HTRF is similar to fluorescence resonance 20 energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely 25 used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & 30 Carson, 1998, *Anal. Biochem.* 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

### 5.5.3. Surface Plasmon Resonance ("SPR")

5 Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the association project in "real time" (Schuck, 1997, *Annu. Rev. Biophys. Biomol. Struct.* 26:541-566).

10 The principle of SPR is summarized by Xavier *et al.* (*Trends Biotechnol.*, 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred 15 nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle 20 at which absorption occurs is very sensitive to the refractive changes in the external medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide 25 composition (the refractive index change is different for proteins and nucleic acids). When the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

30 In the present invention, the target RNA may be immobilized to the sensor surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (*Methods Mol. Biol.*, 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine 35 monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, *J. Biol. Chem.* 274:17452-17463) and the aminoglycoside antibiotics with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, *J. Am. Chem. Soc.* 119:3641-3648; Wong *et al.*, 1998, *Chem. Biol.* 5:397-406).

5 In one embodiment of the present invention, the target nucleic acid can be immobilized to a sensor surface (e.g., by a streptavidin-biotin linkage) and SPR can be used to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

#### 5.5.4. Mass Spectrometry

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for 10 background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a 15 memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each 20 time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and 25 false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry 30 with the eluted analyte (Sonksen *et al.*, 1998, *Anal. Chem.* 70:2731-2736; Nelson & Krone, 1999, *J. Mol. Recog.* 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make SPR combined with mass spectroscopy a high-throughput method for biomolecular 35 interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processes described *supra*. Furthermore, mass spectrometry can also be used to elucidate the structure of the test compound.

#### **5.5.5. Scintillation Proximity Assay (“SPA”)**

10 Scintillation proximity assay (“SPA”) is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, *Drug Discov. Today* 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, *Drug Discov. Today* 4:350-362).

15 Screening for small molecules that inhibit Tat peptide:TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, *Bioorg. Med. Chem.* 5:1173-1184; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212). A similar approach can be used to identify small molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

20 SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, *J. Biomol. Screening* 3:43-48). Thus, one embodiment of the present invention comprises (a) labeling of the target nucleic acid with a radioactive or fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the microplate.

25

#### **5.5.6. Structure-Activity Relationships (“SAR”) by NMR Spectroscopy**

30 NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). The determination of structure-activity relationships (“SAR”) by NMR is the first method for NMR described in which small molecules that

bind adjacent subsites are identified by two-dimentional  $^1\text{H}$ - $^{15}\text{N}$  spectra of the target protein (Shuker *et al.*, 1996, *Science* 274:1531-1534). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed field gradient diffusion measurements (Moore, 1999, *Curr. Opin. Biotechnol.* 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, *Chem. Biol.* 6:755-769).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by 10 NMR can also be used to elucidate the structure of the test compound.

#### **5.5.7. Size Exclusion Chromatography**

In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic acid from a complex mixture of compounds. Size-exclusion chromatography separates molecules based 15 on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and forms a complex array of pores. Separation is accomplished by the inclusion or exclusion of 20 molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target nucleic acid is incubated with a mixture of test compounds while free in solution and 25 allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test compounds (e.g., Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput 30 separation of mixtures (e.g., PLASMID 96-well SEC plates manufactured by Millipore).

#### **5.5.8. Affinity Chromatography**

In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a 35 complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then

captured to a solid support once equilibrium has been established; alternatively, target nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

5 The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

10 Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be 15 accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer, and addition of non-specific competitor to the wash buffer.

20 In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (e.g., in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic acid can be determined by the presence of the detectable label on the test compound using fluorescence.

25 Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the 30 application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent 35 concentration by lyophilization of the eluted compound (e.g., 0 M to 5 M ammonium acetate).

### 5.5.9. Nanoparticle Aggregation

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to  $\text{Ag}_2\text{S}$ ,  $\text{ZnS}$ ,  $\text{CdS}$ ,  $\text{CdTe}$ ,  $\text{Au}$ , or  $\text{TiO}_2$ . Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, *J. Am. Chem. Soc.* 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

## 5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, e.g., by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

25 A number of physico-chemical techniques can be used for the de novo characterization of test compounds bound to the target.

### 5.6.1. Mass Spectrometry

Mass spectrometry (e.g., electrospray ionization (“ESI”) and matrix-assisted laser desorption-ionization (“MALDI”), Fourier-transform ion cyclotron resonance (“FT-ICR”)) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

5 MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, *J. Biol. Chem.* 272:32084-32091).  
10 However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

15 ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, *Anal. Chem.* 69:5130-5135).

20 Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, *Anal. Chem.* 71:3436-3440; Griffey *et al.*, 1999, *Proc. Natl. Acad. Sci. USA* 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

25 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

#### 5.6.2. NMR Spectroscopy

30 As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimentional NMR, two-dimentional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

35 Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

### 5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam.

25

The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, must are observable only with one or

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a 15 sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

20 In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this method, the spectra of compounds with known structure are recorded so that comparison with these spectra can identify compounds again when isolated from RNA binding experiments.

#### 5.6.4. Microwave Spectroscopy

25 In another embodiment, the microwave spectra of a test compound can be used to elucidate the structure of the test compound. For example, as described in U.S. Patent Nos. 6,395,480; 6,376,258; 6,368,795; 6,340,568; 6,338,968; 6,287,874; and 6,287,776 to Hefti, the disclosures of which are hereby incorporated by reference, the unique dielectric properties of molecules and binding complexes, such as hybridization 30 complexes formed between a nucleic acid probe and a nucleic acid target, molecular binding events, and protein/ligand complexes, result in varying microwave spectra which can be measured. The molecule's dielectric properties can be observed by coupling a test signal to the molecule and observing the resulting signal. When the test signal excites the molecule at a frequency within the molecule's dispersion regime, especially at a resonant 35 frequency, the molecule will interact strongly with the signal, and the resulting signal will exhibit dramatic variations in its measured amplitude and phase, thereby generating a

unique signal response. This response can be used to detect and identify the bound molecular structure. In addition, because most molecules will exhibit different dispersion properties over the same or different frequency bands, each generates a unique signal response which can be used to identify the molecular structure.

5

### 5.7. Secondary Biological Screens

The test compounds identified in the nonsense suppression assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using an *in vitro* transcribed RNA from the gene with the premature translation termination codon and *in vitro* translating that RNA in a cell-free translation extract. The activity of the lead compound in the *in vitro* translation mixture can be determined by any method that measures increased expression of the full-length gene, *i.e.*, past the premature termination codon. For example, expression of a functional protein from the full-length gene (*e.g.*, a reporter gene) can be measured to determine the effect of the lead compound on premature translation termination and/or nonsense-mediated mRNA decay in an *in vitro* system.

In addition, the lead compound can be tested in a host cell engineered to contain the RNA with the premature translation termination codon controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the RNA with the premature translation termination codon. Compounds that modulate premature translation termination and/or nonsense-mediated mRNA decay *in vivo* will result in increased expression of the full-length gene, *i.e.*, past the premature termination codon. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA with the premature translation termination codon in the presence and absence of the lead compound. Both the *in vitro* and *in vivo* nonsense suppression assays used herein and as described in International Patent Publication WO 01/44516, which is incorporated by reference in its entirety, can be used to identify lead compounds can also be used to determine an EC<sub>50</sub> for the lead compounds.

The test compounds identified in the nonsense suppression assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the RNA with the premature translation termination codon controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the RNA with the premature translation termination codon. Compounds that modulate premature translation termination and/or nonsense-mediated

5 mRNA decay *in vivo* will result in increased expression of the full-length gene, *i.e.*, past the premature termination codon. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA with the premature translation termination codon in the presence and absence of the lead compound. Both the *in vitro* and *in vivo* nonsense suppression assays used herein and as described in International Patent Publication WO 01/44516, which is incorporated by reference in its entirety, can be used to identify lead compounds can also be used to determine an EC<sub>50</sub> for the lead compounds.

10 Animal model systems can also be used to demonstrate the safety and efficacy of the lead compounds identified in the nonsense suppression assays described above. The lead compounds identified in the nonsense suppression assay can then be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse.

15 Examples of animal models for cystic fibrosis include, but are not limited to, cftr(-/-) mice (see, *e.g.*, Freedman *et al.*, 2001, *Gastroenterology* 121(4):950-7), cftr(tm1HGU/tm1HGU) mice (see, *e.g.*, Bernhard *et al.*, 2001, *Exp Lung Res* 27(4):349-66), CFTR-deficient mice with defective cAMP-mediated Cl(-) conductance (see, *e.g.*, Stotland *et al.*, 2000, *Pediatr Pulmonol* 30(5):413-24), and C57BL/6-  
20 Cftr(m1UNC)/Cftr(m1UNC) knockout mice (see, *e.g.*, Stotland *et al.*, 2000, *Pediatr Pulmonol* 30(5):413-24).

25 Examples of animal models for muscular dystrophy include, but are not limited to, mouse, hamster, cat, dog, and *C. elegans*. Examples of mouse models for muscular dystrophy include, but are not limited to, the dy-/- mouse (see, *e.g.*, Connolly *et al.*, 2002, *J Neuroimmunol* 127(1-2):80-7), a muscular dystrophy with myositis (mdm) mouse mutation (see, *e.g.*, Garvey *et al.*, 2002, *Genomics* 79(2):146-9), the mdx mouse (see, *e.g.*, Nakamura *et al.*, 2001, *Neuromuscul Disord* 11(3):251-9), the utrophin-dystrophin knockout (dko) mouse (see, *e.g.*, Nakamura *et al.*, 2001, *Neuromuscul Disord* 11(3):251-9), the dy/dy mouse (see, *e.g.*, Dubowitz *et al.*, 2000, *Neuromuscul Disord* 10(4-5):292-8), the mdx(Cv3) mouse model (see, *e.g.*, Pillers *et al.*, 1999, *Laryngoscope* 109(8):1310-2), and the myotonic ADR-MDX mutant mice (see, *e.g.*, Kramer *et al.*, 1998, *Neuromuscul Disord* 8(8):542-50). Examples of hamster models for muscular dystrophy include, but are not limited to, sarcoglycan-deficient hamsters (see, *e.g.*, Nakamura *et al.*, 2001, *Am J Physiol Cell Physiol* 281(2):C690-9) and the BIO 14.6 dystrophic hamster (see, *e.g.*, Schlenker & Burbach, 1991, *J Appl Physiol* 71(5):1655-62). An example of a feline model for muscular dystrophy includes, but is not limited to, the hypertrophic feline  
30  
35

5 muscular dystrophy model (see, e.g., Gaschen & Burgunder, 2001, *Acta Neuropathol (Berl)* 101(6):591-600). Canine models for muscular dystrophy include, but are not limited to, golden retriever muscular dystrophy (see, e.g., Fletcher *et al.*, 2001, *Neuromuscul Disord* 11(3):239-43) and canine X-linked muscular dystrophy (see, e.g., Valentine *et al.*, 1992, *Am J Med Genet* 42(3):352-6). Examples of *C. elegans* models for muscular dystrophy are described in Chamberlain & Benian, 2000, *Curr Biol* 10(21):R795-7 and Culette & Sattelle, 2000, *Hum Mol Genet* 9(6):869-77.

10 Examples of animal models for familial hypercholesterolemia include, but are not limited to, mice lacking functional LDL receptor genes (see, e.g., Aji *et al.*, 1997, *Circulation* 95(2):430-7), Yoshida rats (see, e.g., Fantappie *et al.*, 1992, *Life Sci* 50(24):1913-24), the JCR:LA-cp rat (see, e.g., Richardson *et al.*, 1998, *Atherosclerosis* 138(1):135-46), swine (see, e.g., Hasler-Rapacz *et al.*, 1998, *Am J Med Genet* 76(5):379-86), and the Watanabe heritable hyperlipidaemic rabbit (see, e.g., Tsutsumi *et al.*, 2000, *Arzneimittelforschung* 50(2):118-21; Harsch *et al.*, 1998, *Br J Pharmacol* 124(2):227-82; 15 and Tanaka *et al.*, 1995, *Atherosclerosis* 114(1):73-82).

20 An example of an animal model for human cancer in general includes, but is not limited to, spontaneously occurring tumors of companion animals (see, e.g., Vail & MacEwen, 2000, *Cancer Invest* 18(8):781-92). Examples of animal models for lung cancer include, but are not limited to, lung cancer animal models described by Zhang & Roth (1994, *In Vivo* 8(5):755-69) and a transgenic mouse model with disrupted p53 function (see, e.g., Morris *et al.*, 1998, *J La State Med Soc* 150(4):179-85). An example of an animal model for breast cancer includes, but is not limited to, a transgenic mouse that overexpresses cyclin D1 (see, e.g., Hosokawa *et al.*, 2001, *Transgenic Res* 10(5):471-8). 25 An example of an animal model for colon cancer includes, but is not limited to, a TCRbeta and p53 double knockout mouse (see, e.g., Kado *et al.*, 2001, *Cancer Res* 61(6):2395-8). Examples of animal models for pancreatic cancer include, but are not limited to, a metastatic model of Panc02 murine pancreatic adenocarcinoma (see, e.g., Wang *et al.*, 2001, *Int J Pancreatol* 29(1):37-46) and nu-nu mice generated in subcutaneous pancreatic tumours (see, e.g., Ghaneh *et al.*, 2001, *Gene Ther* 8(3):199-208). Examples of animal models for 30 non-Hodgkin's lymphoma include, but are not limited to, a severe combined immunodeficiency ("SCID") mouse (see, e.g., Bryant *et al.*, 2000, *Lab Invest* 80(4):553-73) and an IgHmu-HOX11 transgenic mouse (see, e.g., Hough *et al.*, 1998, *Proc Natl Acad Sci USA* 95(23):13853-8). An example of an animal model for esophageal cancer includes, but is not limited to, a mouse transgenic for the human papillomavirus type 16 E7 oncogene 35 (see, e.g., Herber *et al.*, 1996, *J Virol* 70(3):1873-81). Examples of animal models for

colorectal carcinomas include, but are not limited to, Apc mouse models (see, e.g., Fodde & Smits, 2001, Trends Mol Med 7(8):369-73 and Kuraguchi *et al.*, 2000, Oncogene 19(50):5755-63). An example of an animal model for neurofibromatosis includes, but is not limited to, mutant NF1 mice (see, e.g., Cichowski *et al.*, 1996, Semin Cancer Biol 7(5):291-8). Examples of animal models for retinoblastoma include, but are not limited to, transgenic mice that express the simian virus 40 T antigen in the retina (see, e.g., Howes *et al.*, 1994, Invest Ophthalmol Vis Sci 35(2):342-51 and Windle *et al.*, 1990, Nature 343(6259):665-9) and inbred rats (see, e.g., Nishida *et al.*, 1981, Curr Eye Res 1(1):53-5 and Kobayashi *et al.*, 1982, Acta Neuropathol (Berl) 57(2-3):203-8). Examples of animal models for Wilm's tumor include, but are not limited to, a WT1 knockout mice (see, e.g., Scharnhorst *et al.*, 1997, Cell Growth Differ 8(2):133-43), a rat subline with a high incidence of nephroblastoma (see, e.g., Mesfin & Breech, 1996, Lab Anim Sci 46(3):321-6), and a Wistar/Furth rat with Wilms' tumor (see, e.g., Murphy *et al.*, 1987, Anticancer Res 7(4B):717-9).

Examples of animal models for retinitis pigmentosa include, but are not limited to, the Royal College of Surgeons ("RCS") rat (see, e.g., Vollrath *et al.*, 2001, Proc Natl Acad Sci USA 98(22):12584-9 and Hanitzsch *et al.*, 1998, Acta Anat (Basel) 162(2-3):119-26), a rhodopsin knockout mouse (see, e.g., Jaissle *et al.*, 2001, Invest Ophthalmol Vis Sci 42(2):506-13), and Wag/Rij rats (see, e.g., Lai *et al.*, 1980, Am J Pathol 98(1):281-4).

Examples of animal models for cirrhosis include, but are not limited to, CCl<sub>4</sub>-exposed rats (see, e.g., Kloehn *et al.*, 2001, Horm Metab Res 33(7):394-401) and rodent models instigated by bacterial cell components or colitis (see, e.g., Vierling, 2001, Best Pract Res Clin Gastroenterol 15(4):591-610).

Examples of animal models for hemophilia include, but are not limited to, rodent models for hemophilia A (see, e.g., Reipert *et al.*, 2000, Thromb Haemost 84(5):826-32; Jarvis *et al.*, 1996, Thromb Haemost 75(2):318-25; and Bi *et al.*, 1995, Nat Genet 10(1):119-21), canine models for hemophilia A (see, e.g., Gallo-Penn *et al.*, 1999, Hum Gene Ther 10(11):1791-802 and Connelly *et al.*, 1998, Blood 91(9):3273-81), murine models for hemophilia B (see, e.g., Snyder *et al.*, 1999, Nat Med 5(1):64-70; Wang *et al.*, 1997, Proc Natl Acad Sci USA 94(21):11563-6; and Fang *et al.*, 1996, Gene Ther 3(3):217-22), canine models for hemophilia B (see, e.g., Mount *et al.*, 2002, Blood 99(8):2670-6; Snyder *et al.*, 1999, Nat Med 5(1):64-70; Fang *et al.*, 1996, Gene Ther 3(3):217-22); and Kay *et al.*, 1994, Proc Natl Acad Sci USA 91(6):2353-7), and a rhesus macaque model for hemophilia B (see, e.g., Lozier *et al.*, 1999, Blood 93(6):1875-81).

Examples of animal models for von Willebrand disease include, but are not limited to, an inbred mouse strain RIIS/J (see, e.g., Nichols *et al.*, 1994, 83(11):3225-31 and Sweeney *et al.*, 1990, 76(11):2258-65), rats injected with botrocetin (see, e.g., Sanders *et al.*, 1988, Lab Invest 59(4):443-52), and porcine models for von Willebrand disease (see, e.g., Nichols *et al.*, 1995, Proc Natl Acad Sci USA 92(7):2455-9; Johnson & Bowie, 1992, J Lab Clin Med 120(4):553-8); and Brinkhous *et al.*, 1991, Mayo Clin Proc 66(7):733-42).

Examples of animal models for  $\beta$ -thalassemia include, but are not limited to, murine models with mutations in globin genes (see, e.g., Lewis *et al.*, 1998, Blood 91(6):2152-6; Raja *et al.*, 1994, Br J Haematol 86(1):156-62; Popp *et al.*, 1985, 445:432-44; and Skow *et al.*, 1983, Cell 34(3):1043-52).

Examples of animal models for kidney stones include, but are not limited to, genetic hypercalciuric rats (see, e.g., Bushinsky *et al.*, 1999, Kidney Int 55(1):234-43 and Bushinsky *et al.*, 1995, Kidney Int 48(6):1705-13), chemically treated rats (see, e.g., Grases *et al.*, 1998, Scand J Urol Nephrol 32(4):261-5; Burgess *et al.*, 1995, Urol Res 23(4):239-42; Kumar *et al.*, 1991, J Urol 146(5):1384-9; Okada *et al.*, 1985, Hinyokika Kiyo 31(4):565-77; and Bluestone *et al.*, 1975, Lab Invest 33(3):273-9), hyperoxaluric rats (see, e.g., Jones *et al.*, 1991, J Urol 145(4):868-74), pigs with unilateral retrograde flexible nephroscopy (see, e.g., Seifmah *et al.*, 2001, 57(4):832-6), and rabbits with an obstructed upper urinary tract (see, e.g., Itatani *et al.*, 1979, Invest Urol 17(3):234-40).

Examples of animal models for ataxia-telangiectasia include, but are not limited to, murine models of ataxia-telangiectasia (see, e.g., Barlow *et al.*, 1999, Proc Natl Acad Sci USA 96(17):9915-9 and Inoue *et al.*, 1986, Cancer Res 46(8):3979-82).

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (cf. Lamothe, *et al.* 1997, J. Med. Chem. 40: 3542; Mottola *et al.* 1996, J. Med. Chem. 39: 285; Beusen *et al.* 1995, Biopolymers 36: 181; P. Fossa *et al.* 1998, Comput. Aided Mol. Des. 12: 361), QSAR development (cf. Siddiqui *et al.* 1999, J. Med. Chem. 42: 4122; Barreca *et al.* 1999 Bioorg. Med. Chem. 7: 2283; Kroemer *et al.* 1995, J. Med. Chem. 38: 4917; Schaal *et al.* 2001, J. Med. Chem. 44: 155; Buolamwini & Assefa 2002, J. Mol. Chem. 45: 84), Virtual docking and screening/scoring (cf. Anzini *et al.* 2001, J. Med. Chem. 44: 1134; Faaland *et al.* 2000, Biochem. Cell. Biol. 78: 415; Silvestri *et al.* 2000, Bioorg.

Med. Chem. 8: 2305; J. Lee *et al.* 2001, Bioorg. Med. Chem. 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAMotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

5 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinaly & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (e.g. RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the 10 calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (e.g., Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to 15 develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

20

25

30

### 5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is associated with premature translation termination and/or nonsense-mediated mRNA decay. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with premature translation termination and/or nonsense-mediated mRNA decay.

In a preferred embodiment, it is first determined that the patient is suffering from a disease associated with premature translation termination and/or nonsense-mediated mRNA decay. In a preferred embodiment, the DNA of the patient can be sequenced or subject to Southern blot, polymerase chain reaction ("PCR"), use of the Short Tandem Repeat ("STR"), or polymorphic length restriction fragments ("RFLP") analysis to determine if a nonsense mutation is present in the DNA of the patient. Alternatively, it can be determined if altered levels of the protein with the nonsense mutation are expressed in the patient by western blot or other immunoassays. Such methods are well known to one of skill in the art.

In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with a disease caused and/or associated with nonsense suppression. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

In a preferred embodiment, the compounds identified using the methods of the present invention are used to treat or prevent a disease caused by one or more nonsense mutations. Examples of diseases caused by nonsense mutations include, but are not limited to, cystic fibrosis, muscular dystrophy, heart disease, lung cancer, breast cancer, colon cancer, pancreatic cancer, non-Hodgkin's lymphoma, ovarian cancer, esophageal cancer, colorectal carcinomas, neurofibromatosis, retinoblastoma, Wilm's tumor, retinitis pigmentosa, collagen disorders, cirrhosis, Tay-Sachs disease, blood disorders, kidney stones, and ataxia-telangiectasia. Genes that contain one or more nonsense mutations that are potentially involved in causing disease are presented in table form according to 10 chromosome location in Example 9 *infra*.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus 15 injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, 20 capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, 25 sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and 30 not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

In certain embodiments, it may be desirable to introduce the compound or a 35 pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular

injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

5 Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

10 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

15 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, e.g., Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533 may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507; Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see 20 *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science* 228:190; During *et al.*, 1989, *Ann. Neurol.* 25:351; Howard *et al.*, 1989, *J. Neurosurg.* 71:105). In yet another embodiment, a controlled-release system can be placed in proximity 25 of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

30 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

35 In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient,

or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see e.g., U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions.

In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise 5 sterile isotonic aqueous buffer. Where necessary, the compositions may also include a 10 solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at 15 the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle 20 containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that 25 will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each 30 patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 35 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer

to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

5 Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal 10 administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

15 Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived 20 from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

25 The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

30 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as  $\alpha$ -amino phosphonic acids and  $\alpha$ -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, 35 lactose and galactose.

**6. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS**

The results presented in this Example indicate that gel mobility shift assays  
 5 can be used to detect the binding of small molecules, such as the Tat peptide and gentamicin, to their respective target RNAs.

**6.1. Materials and Methods**

10

**6.1.1. Buffers**

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl<sub>2</sub>. Tris-borate-EDTA (TBE) buffer is composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl<sub>2</sub>.

**6.1.2. Gel retardation analysis**

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 1); Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCAGAUCUGAGCCUGGGAGCUCUCUGCC-3' (SEQ ID NO: 2); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was 3' labeled with 5'-<sup>32</sup>P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim). For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl<sub>2</sub> and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 3; single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20  $\mu$ l reaction volumes 50 pmoles of <sup>32</sup>P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat<sub>47-58</sub>) in TK or TKM buffer were heated at 90°C 35 for 2 minutes and allow to cool to room temperature (approximately 24 °C) over 2 hours.

Then 10  $\mu$ l of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4 °C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).  
5

### 6.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, *Nature* 327:389-394; Woodcock *et al.*, 1991, *EMBO J.* 10:3099-3103; Yoshizawa *et al.*, 1998, *EMBO J.* 17:6437-6448). Here the use of gel mobility shift assays to monitor RNA-small molecule interactions are described. This approach allows for rapid visualization of small molecule-RNA interactions based on the difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, *Nature* 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug 20 discovery.  
10  
15  
20

### 6.3. Results

A gel retardation assay was performed using the Tat<sub>47-58</sub> peptide and the TAR RNA oligonucleotide. As shown in Figure 2, in the presence of the Tat peptide, a clear shift 25 is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093).

Based on the results of Figure 2, it was hypothesized that RNA interactions 30 with small organic molecules could also be visualized using this method. As shown in Figure 3, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a 35 defined structure in solution can be monitored using this approach. In addition, as shown in Figure 3, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 3, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 4, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 4 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10  $\mu$ g/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg- and temperature dependent. As shown in Figure 5, when  $MgCl_2$  is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

25 **7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA  
BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS  
BY CAPILLARY ELECTROPHORESIS**

---

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

**7.1. Materials and Methods**

**7.1.1. Buffers**

35 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM  $MgCl_2$ . Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl<sub>2</sub>.

5

### **7.1.2. Gel Retardation Analysis Using Capillary Electrophoresis**

RNA oligonucleotides were purchased from Dhamacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCAGAUCUGAGGCCUGGGAGCUCUCUGCC-3' (SEQ ID NO: 2); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl<sub>2</sub> and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 3); single letter amino acid code). In 20  $\mu$ l reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat<sub>47-58</sub>) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

20

### **7.2. Results**

As presented in the previous Example in Section 6, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat<sub>47-58</sub> peptide and the TAR RNA oligonucleotide. As shown in Figure 6 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093).

35

**8. EXAMPLE: COMPOUNDS THAT MODULATE TRANSLATION TERMINATION BIND SPECIFIC REGIONS OF 28S rRNA**

Data is presented in this Example that demonstrates that specific regions of the 28S rRNA are involved in modulating translation termination in mammalian cells. Compounds that interact in these regions or modulate local changes within these regions of the ribosome (e.g., alter base pairing interactions, base modification or modulate binding of trans-acting factors that bind to these regions) have the potential to modulate translation termination. These regions are conserved from prokaryotes to eukaryotes, but the role of these regions in modulating translation termination has not been realized in eukaryotes. In bacteria, when a short RNA fragment, complementary to the *E. coli* 23S rRNA segment comprising nucleotides 735 to 766 (in domain II), is expressed *in vivo*, suppression of UGA nonsense mutations, but not UAA or UAG, results (Chernyaeva *et al.*, 1999, J Bacteriol 181:5257-5262). Other regions of the 23S rRNA in *E. coli* have been implicated in nonsense suppression including the GTPase center in domain II (nt 1034-1120; Jemiolo *et al.*, 1995, Proc. Nat. Acad. Sci. 92:12309-12313).

**8.1. Materials and Methods**

**8.1.1. Small Molecules Involved in Modulating Translation Termination**

Small molecules involved in modulating translation termination, *i.e.*, nonsense suppression, were used in the footprinting experiments presented in Figures 2 to 6 and are listed as Compound A (molecular formula  $C_{19}H_{21}NO_4$ ), Compound B (molecular formula  $C_{19}H_{21}N_2O_5$ ), Compound C (molecular formula  $C_{12}H_{15}N_5O$ ), Compound D (molecular formula  $C_{23}H_{15}O_3Br$ ), Compound E (molecular formula  $C_{19}H_{21}NO_4$ ), Compound F, Compound G (molecular formula  $C_{12}H_{15}N_5O$ ), Compound H (molecular formula  $C_{23}H_{15}NO_5$ ), Compound I (molecular formula  $C_{23}H_{15}NO_5$ ), Compound J, and Compound K.

**8.1.2. Preparation of a Translation Extract from HeLa cells**

HeLa S3 cells were grown to a density of  $10^6$  cells/ml in DMEM; 5%CO<sub>2</sub>, 10%FBS, 1X P/S in a spinner flask. Cells were harvested by spinning at 1000Xg. Cells were washed twice with phosphate buffered saline. The cell pellet was on ice for 12-24 hours before proceeding. By letting the cells sit on ice, the activity of the extract is increased up to twenty-fold. The length of time on ice can range from 0 hours to 1 week. The cells were resuspended in 1.5 volumes (packed cell volume) of hypotonic buffer (10

mM HEPES (KOH) pH 7.4; 15 mM KCl; 1.5 mM Mg(OAc)<sub>2</sub>; 0.5 mM Pefabloc (Roche); 2 mM DTT). The cells were allowed to swell for 5 minutes on ice, dounce homogenized with 10 to 100 strokes using a tight-fitting pestle, and spun for 10 minutes at 12000Xg at 4°C in a Sorvall SS-34 rotor. The supernatant was collected with a Pasteur pipet without disturbing the lipid layer, transferred into Eppendorf tubes (50 to 200 µl aliquots), and immediately frozen in liquid nitrogen.

### 8.1.3. Footprinting

Ribosomes prepared from HeLa cells were incubated with the small molecules (at a concentration of 100 µM), followed by treatment with chemical modifying agents (dimethyl sulfate [DMS] and kethoxal [KE]). Following chemical modification, rRNA was phenol-chloroform extracted, ethanol precipitated, analyzed in primer extension reactions using end-labeled oligonucleotides hybridizing to different regions of the rRNAs and resolved on 6% polyacrylamide gels. The probes used for primer extension cover the entire 18S (7 oligonucleotide primers), 28S (24 oligonucleotide primers), and 5S (one primer) rRNAs are presented in Table 1 (also see, e.g., Gonzalez et al., 1985 Proc Natl Acad Sci U S A. 82(22):7666-70 and McCallum & Maden, 1985, Biochem. J. 232 (3): 725-733). Controls in these experiments include DMSO (a control for changes in rRNA accessibility induced by DMSO), paromomycin (a marker for 18S rRNA binding), and anisomycin (a marker for 28S rRNA binding).

Table 1: 18S, 28S, and 5S rRNA primers

5S#1	AAAGCCTACAGCACCC	SEQ ID NO.: 4
25 28S#1	TACTGAGGGAATCCTGG	SEQ ID NO.: 5
28S#2	TTACCACCCGCTTGGG	SEQ ID NO.: 6
28S#3	GGGGGCGGGAAAGATCC	SEQ ID NO.: 7
28S#4	CCCCGAGCCACCTTCCC	SEQ ID NO.: 8
30 28S#5	GGCCCCGGGATTGGCG	SEQ ID NO.: 9
28S#6	CACTGGGGACAGTCCGC	SEQ ID NO.: 10
28S#7	CGCGGCGGGCGAGACGGG	SEQ ID NO.: 11
28S#8	GAGGGAAACTCGGAGGG	SEQ ID NO.: 12
35 28S#9	CATCGGGCGCCTAACCC	SEQ ID NO.: 13

28S#10	CGACGCACACCACACGC	SEQ ID NO.: 14
28S#11	CCAAGATCTGCACCTGC	SEQ ID NO.: 15
28S#12	TTACCGCACTGGACGCC	SEQ ID NO.: 16
5 28S#13	GCCAGAGGCTGTTACC	SEQ ID NO.: 17
28S#14	TGGGGAGGGAGCGAGCGGCC	SEQ ID NO.: 18
28S#15	AAGGGCCCGGCTCGCGTCC	SEQ ID NO.: 19
28S#16	AGGGCGGGGGGACGAACCGC	SEQ ID NO.: 20
10 28S#17	TTAACACAGTCGGATTCCCCTGG	SEQ ID NO.: 21
28S#18	TTCATCCATTATGCGCG	SEQ ID NO.: 22
28S#19	AGTAGTGGTATTCACCGG	SEQ ID NO.: 23
28S#20	ACGGGAGGTTCTGTCC	SEQ ID NO.: 24
15 28S#21	ACAATGATAGGAAGAGCCG	SEQ ID NO.: 25
28S#22	AGGCAGTCAGTCATAATCCC	SEQ ID NO.: 26
28S#23	TCCGCACCGGACCCGGTCC	SEQ ID NO.: 27
28S#24	GGGCTAGTTGATTCCGGCAGGTGAGTTG	SEQ ID NO.: 28
20 18S#1	TCTCCGGAATCGAACCCCT	SEQ ID NO.: 29
18S#2	ATT ACC GCGGCTGCTGGC	SEQ ID NO.: 30
18S#3	TTGGCAAATGCTTCGC	SEQ ID NO.: 31
18S#4	CCGTCAATTCTTTAACGTTTC	SEQ ID NO.: 32
25 18S#5	AGGGCATCACAGACCTGTTAT	SEQ ID NO.: 33
18S#6	CGACGGGCGGTGTAC	SEQ ID NO.: 34
18S#7	CCGCAGGTTCACCTACGG	SEQ ID NO.: 35

### 8.2. Results

30 The results of these foot-printing experiments (see, *e.g.*, Figures 7 to 11) indicated that the small molecules involved in modulating translation termination alter the accessibility of the chemical modifying agents to specific nucleotides in the 28S rRNA. More specifically, the regions protected by the small molecules include a conserved region in the vicinity of the peptidyl transferase center (domain V, see, *e.g.*, Figures 7 and 8) 35 implicated in peptide bond formation and a conserved region in domain II (see, *e.g.*, Figures

9, 10, and 11) that may interact with the peptidyl transferase center based on binding of vernamycin B to both these areas (Vannuffel et al., 1994, Nucleic Acids Res. 22(21):4449-53).

5

**9. EXAMPLE: HUMAN DISEASE GENES SORTED BY CHROMOSOME**

Table 2: Genes, Locations and Genetic Disorders on Chromosome 1

Gene	GDB Accession ID	OMIM Link
ABCA4	GDB:370748	MACULAR DEGENERATION, SENILE STARGARDT DISEASE 1; STGD1 ATP BINDING CASSETTE TRANSPORTER; ABCR RETINITIS PIGMENTOSA-19; RP19
ABCD3	GDB:131485	PEROXISOMAL MEMBRANE PROTEIN 1; PXMP1
ACADM	GDB:118958	ACYL-CoA DEHYDROGENASE, MEDIUM-CHAIN; ACADM
AGL	GDB:132644	GLYCOGEN STORAGE DISEASE III
AGT	GDB:118750	ANGIOTENSIN I; AGT
ALDH4A1	GDB:9958827	HYPERPROLINEMIA, TYPE II
ALPL	GDB:118730	PHOSPHATASE, LIVER ALKALINE; ALPL HYPOPHOSPHATASIA, INFANTILE
AMPD1	GDB:119677	ADENOSINE MONOPHOSPHATE DEAMINASE-1; AMPD1
APOA2	GDB:119685	APOLIPOPROTEIN A-II; APOA2
AVSD1	GDB:265302	ATRIOVENTRICULAR SEPTAL DEFECT; AVSD
BRCD2	GDB:9955322	BREAST CANCER, DUCTAL, 2; BRCD2
C1QA	GDB:119042	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, ALPHA POLYPEPTIDE; C1QA
C1QB	GDB:119043	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, BETA POLYPEPTIDE; C1QB

35

Gene	GDB Accession ID	OMIM Link
5	C1QG	COMPLEMENT COMPONENT 1, q SUBCOMPONENT, GAMMA POLYPEPTIDE; C1QG
	C8A	COMPLEMENT COMPONENT-8, DEFICIENCY OF
	C8B	COMPLEMENT COMPONENT-8, DEFICIENCY OF, TYPE II
10	CACNA1S	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, L TYPE, ALPHA 1S SUBUNIT; CACNA1S PERIODIC PARALYSIS I MALIGNANT HYPERTHERMIA SUSCEPTIBILITY-5; MHS5
	CCV	CATARACT, CONGENITAL, VOLKMANN TYPE; CCV
15	CD3Z	CD3Z ANTIGEN, ZETA POLYPEPTIDE; CD3Z
	CDC2L1	PROTEIN KINASE p58; PK58
	CHML	CHOROIDEREMIA-LIKE; CHML
20	CHS1	CHEDIAK-HIGASHI SYNDROME; CHS1
	CIAS1	COLD HYPERSENSITIVITY URTICARIA, DEAFNESS, AND AMYLOIDOSIS
25	CLCNKB	CHLORIDE CHANNEL, KIDNEY, B; CLCNKB
	CMD1A	CARDIOMYOPATHY, DILATED 1A; CMD1A
	CMH2	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 2; CMH2
	CMM	MELANOMA, MALIGNANT
	COL11A1	COLLAGEN, TYPE XI, ALPHA-1; COL11A1
35		

Gene	GDB Accession ID	OMIM Link
COL9A2	GDB:138310	COLLAGEN, TYPE IX, ALPHA-2 CHAIN; COL9A2 EPIPHYSEAL DYSPLASIA, MULTIPLE, 2; EDM2
CPT2	GDB:127272	MYOPATHY WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE II HYPOGLYCEMIA, HYPOKETOTIC, WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE CARNITINE PALMITOYLTRANSFERASE II; CPT2
CRB1	GDB:333930	RETINITIS PIGMENTOSA-12; RP12
<hr/>		
CSE	GDB:596182	CHOREOATHETOSIS/SPASTICITY, EPISODIC; CSE
CSF3R	GDB:126430	COLONY STIMULATING FACTOR 3 RECEPTOR, GRANULOCYTE; CSF3R
CTPA	GDB:9863168	CATARACT, POSTERIOR POLAR
CTSK	GDB:453910	PYCNODYSOSTOSIS CATHEPSIN K; CTSK
DBT	GDB:118784	MAPLE SYRUP URINE DISEASE, TYPE 2
DIO1	GDB:136449	THYROXINE DEIODINASE TYPE I; TXDI1
DISC1	GDB:9992707	DISORDER-2; SCZD2
<hr/>		
DPYD	GDB:364102	DIHYDROPRYIMIDINE DEHYDROGENASE; DPYD
EKV	GDB:119106	ERYTHROKERATODERMIA VARIABILIS; EKV
ENO1	GDB:119871	PHOSPHOPYRUVATE HYDRATASE; PPH
ENO1P	GDB:135006	PHOSPHOPYRUVATE HYDRATASE; PPH

Gene	GDB Accession ID	OMIM Link
EPB41	GDB:119865	ERYTHROCYTE MEMBRANE PROTEIN BAND 4.1; EPB41 HEREDITARY HEMOLYTIC
5 EPHX1	GDB:119876	EPOXIDE HYDROLASE 1, MICROSOMAL; EPHX1
F13B	GDB:119893	FACTOR XIII, B SUBUNIT; F13B
10 F5	GDB:119896	FACTOR V DEFICIENCY
FCGR2A	GDB:119903	Fc FRAGMENT OF IgG, LOW AFFINITY IIa, RECEPTOR FOR; FCGR2A
FCGR2B	GDB:128183	Fc FRAGMENT OF IgG, LOW AFFINITY IIa, RECEPTOR FOR; FCGR2A
15 FCGR3A	GDB:119904	Fc FRAGMENT OF IgG, LOW AFFINITY IIIa, RECEPTOR FOR; FCGR3A
FCHL	GDB:9837503	HYPERLIPIDEMIA, COMBINED
<hr/>		
20 FH	GDB:119133	FUMARATE HYDRATASE; FH LEIOMYOMATA, HEREDITARY MULTIPLE, OF SKIN
FMO3	GDB:135136	FLAVIN-CONTAINING MONOOXYGENASE 3; FMO3 TRIMETHYLAMINURIA
25 FMO4	GDB:127981	FLAVIN-CONTAINING MONOOXYGENASE 2; FMO2
FUCA1	GDB:119237	FUCOSIDOSIS
FY	GDB:119242	BLOOD GROUP--DUFFY SYSTEM; Fy
GALE	GDB:119245	GALACTOSE EPIMERASE DEFICIENCY
30 GBA	GDB:119262	GAUCHER DISEASE, TYPE I; GD I
<hr/>		
GFND	GDB:9958222	GLOMERULAR NEPHRITIS, FAMILIAL, WITH FIBRONECTIN DEPOSITS

Gene	GDB Accession ID	OMIM Link
GJA8	GDB:696369	CATARACT, ZONULAR PULVERULENT 1; CZP1 GAP JUNCTION PROTEIN, ALPHA-8, 50-KD; GJA8
GJB3	GDB:127820	ERYTHROKERATODERMA VARIABILIS; EKV DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 2; DFNA2
10		
GLC3B	GDB:3801939	GLAUCOMA 3, PRIMARY INFANTILE, B; GLC3B
HF1	GDB:120041	H FACTOR 1; HF1
HMGCL	GDB:138445	HYDROXYMETHYLGLUTARICACIDURIA; HMGCL
HPC1	GDB:5215209	PROSTATE CANCER; PRCA1 PROSTATE CANCER, HEREDITARY 1
HRD	GDB:9862254	HYPOPARATHYROIDISM WITH SHORT STATURE, MENTAL RETARDATION, AND SEIZURES
HRPT2	GDB:125253	HYPERPARATHYROIDISM, FAMILIAL PRIMARY, WITH MULTIPLE OSSIFYING JAW
HSD3B2	GDB:134044	ADRENAL HYPERPLASIA II
HSPG2	GDB:126372	HEPARAN SULFATE PROTEOGLYCAN OF BASEMENT MEMBRANE; HSPG2 MYOTONIC MYOPATHY, DWARFISM, CHONDRODYSTROPHY, AND OCULAR AND FACIAL
KCNQ4	GDB:439046	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 2; DFNA2
30		
KCS	GDB:9848740	KENNY-CAFFEY SYNDROME, RECESSIVE FORM
KIF1B	GDB:128645	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, A; CMT2A
35		

Gene	GDB Accession ID	OMIM Link
LAMB3	GDB:251820	LAMININ, BETA 3; LAMB3
LAMC2	GDB:136225	LAMININ, GAMMA 2; LAMC2 EPIDERMOLYSIS BULLOSA LETALIS
LGMD1B	GDB:231606	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 1B; LGMD1B
LMNA	GDB:132146	LAMIN A/C; LMNA LIPODYSTROPHY, FAMILIAL PARTIAL, DUNNIGAN TYPE; LDP1
LOR	GDB:132049	LORICRIN; LOR
MCKD1	GDB:9859381	POLYCYSTIC KIDNEYS, MEDULLARY TYPE
MCL1	GDB:139137	MYELOID CELL LEUKEMIA 1; MCL1
MPZ	GDB:125266	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS MYELIN PROTEIN ZERO; MPZ
MTHFR	GDB:370882	5,10-@METHYLENETETRAHYDROFO LATE REDUCTASE; MTHFR
MTR	GDB:119440	METHYLtetrahydrofolate:L-HO MOCYSTEINE S-METHYLTRANSFERASE; MTR
MUTYH	GDB:9315115	ADENOMATOUS POLYPOSIS OF THE COLON; APC
MYOC	GDB:5584221	GLAUCOMA 1, OPEN ANGLE; GLC1A MYOCILIN; MYOC
NB	GDB:9958705	NEUROBLASTOMA; NB
NCF2	GDB:120223	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-POSITIVE FORM
NEM1	GDB:127387	NEMALINE MYOPATHY 1, AUTOSOMAL DOMINANT; NEM1
NPHS2	GDB:9955617	ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 2; ARVD2

	Gene	GDB Accession ID	OMIM Link
5	NPPA	GDB:118727	NATRIURETIC PEPTIDE PRECURSOR A; NPPA
	NRAS	GDB:119457	ONCOGENE NRAS; NRAS; NRAS1
10	NTRK1	GDB:127897	ONCOGENE TRK NEUROTROPHIC TYROSINE KINASE, RECEPTOR, TYPE 1; NTRK1 NEUROPATHY, CONGENITAL SENSORY, WITH ANHIDROSIS
	OPTA2	GDB:9955577	OSTEOPETROSIS, AUTOSOMAL DOMINANT, TYPE II; OPA2
	PBX1	GDB:125351	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1; PBX1
15	PCHC	GDB:9955586	PHEOCHROMOCYTOMA
	PGD	GDB:119486	6-@PHOSPHOGLUCONATE DEHYDROGENASE, ERYTHROCYTE
	PHA2A	GDB:9955628	PSEUDOHYPOALDOSTERONISM, TYPE II; PHA2
20	PHGDH	GDB:9958261	3-@PHOSPHOGLYCERATE DEHYDROGENASE DEFICIENCY
	PKLR	GDB:120294	PYRUVATE KINASE DEFICIENCY OF ERYTHROCYTE
	PKP1	GDB:4249598	PLAKOPHILIN 1; PKP1
25	PLA2G2A	GDB:120296	PHOSPHOLIPASE A2, GROUP IIA; PLA2G2A
	PLOD	GDB:127821	PROCOLLAGEN-LYSINE, 2-OXOGLUTARATE 5-DIOXYGENASE; PLOD EHRLERS-DANLOS SYNDROME, TYPE VI; E-D VI; EDS VI
30	PPOX	GDB:118852	PROTOPORPHYRINOGEN OXIDASE; PPOX
	PPT	GDB:125227	CEROID-LIPOFUSCINOSIS, NEURONAL 1, INFANTILE; CLN1 PALMITOYL-PROTEIN THIOESTERASE; PPT

	Gene	GDB Accession ID	OMIM Link
	PRCC	GDB:3888215	PAPILLARY RENAL CELL CARCINOMA; PRCC
5	PRG4	GDB:9955719	ARTHROPATHY-CAMPTODACTYLY SYNDROME
10	PSEN2	GDB:633044	ALZHEIMER DISEASE, FAMILIAL, TYPE 4; AD4
	PTOS1	GDB:6279920	PTOSIS, HEREDITARY CONGENITAL 1; PTOS1
	REN	GDB:120345	RENIN; REN
	RFX5	GDB:6288464	REGULATORY FACTOR 5; RFX5
15	RHD	GDB:119551	RHESUS BLOOD GROUP, D ANTIGEN; RHD
	RMD1	GDB:448902	RIPPLING MUSCLE DISEASE-1; RMD1
	RPE65	GDB:226519	RETINAL PIGMENT EPITHELIUM-SPECIFIC PROTEIN, 65-KD; RPE65 AMAUROSIS CONGENITA OF LEBER II
20	SCCD	GDB:9955558	CORNEAL DYSTROPHY, CRYSTALLINE, OF SCHNYDER
	SERPINC1	GDB:119024	ANTITHROMBIN III DEFICIENCY
	SJS1	GDB:1381631	MYOTONIC MYOPATHY, DWARFISM, CHONDRODYSTROPHY, AND OCULAR AND FACIAL
25	SLC19A2	GDB:9837779	THIAMINE-RESPONSIVE MEGLABLASTIC ANEMIA SYNDROME
	SLC2A1	GDB:120627	SOLUTE CARRIER FAMILY 2, MEMBER 1; SLC2A1
30	SPTA1	GDB:119601	ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE HEREDITARY HEMOLYTIC SPECTRIN, ALPHA, ERYTHROCYTIC 1; SPTA1
35	TAL1	GDB:120759	T-CELL ACUTE LYMPHOCYTIC LEUKEMIA 1; TAL1

Gene	GDB Accession ID	OMIM Link
TNFSF6	GDB:422178	APOPTOSIS ANTIGEN LIGAND 1; APT1LG1
5 TNNT2	GDB:221879	TROPONIN-T2, CARDIAC; TNNT2
TPM3	GDB:127872	ONCOGENE TRK TROPOMYOSIN 3; TPM3
TSHB	GDB:120467	THYROID-STIMULATING HORMONE, BETA CHAIN; TSHB
10 UMPK	GDB:120481	URIDINE MONOPHOSPHATE KINASE; UMPK
UOX	GDB:127539	URATE OXIDASE; UOX
UROD	GDB:119628	PORPHYRIA CUTANEA TARDA; PCT
15 USH2A	GDB:120483	USHER SYNDROME, TYPE II; USH2
VMGLOM	GDB:9958134	GLOMUS TUMORS, MULTIPLE
VWS	GDB:120532	CLEFT LIP AND/OR PALATE WITH MUCOUS CYSTS OF LOWER LIP
20 WS2B	GDB:407579	WAARDENBURG SYNDROME, TYPE 2B; WS2B

Table 3: Genes, Locations and Genetic Disorders on Chromosome 2

Gene	GDB Accession ID	Location	OMIM Link
25 ABCB11	GDB:9864786	2q24-2q24 2q24.3-2q24.3	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATIC 2; PFIC2
ABCG5	GDB:10450298	2p21-2p21	PHYTOSTEROLEMIA
ABCG8	GDB:10450300	2p21-2p21	PHYTOSTEROLEMIA
30 ACADL	GDB:118745	2q34-2q35	ACYL-CoA DEHYDROGENASE, LONG-CHAIN, DEFICIENCY OF
ACP1	GDB:118962	2p25-2p25	PHOSPHATASE, ACID, OF ERYTHROCYTE; ACP1
AGXT	GDB:127113	2q37.3-2q37.3	OXALOSIS I
35 AHHR	GDB:118984	2pter-2q31	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 1; CYP1A1

Gene	GDB Accession ID	Location	OMIM Link
ALMS1	GDB:9865539	2p13-2p12 2p14-2p13 2p13.1-2p13.1	ALSTROM SYNDROME
5 ALPP	GDB:119672	2q37.1-2q37.1	ALKALINE PHOSPHATASE, PLACENTAL; ALPP
ALS2	GDB:135696	2q33-2q35	AMYOTROPHIC LATERAL SCLEROSIS 2, JUVENILE; ALS2
10 APOB	GDB:119686	2p24-2p23 2p24-2p24	APOLIPOPROTEIN B; APOB
BDE	GDB:9955730	2q37-2q37	BRACHYDACTYLY, TYPE E; BDE
15 BDMR	GDB:533064	2q37-2q37	BRACHYDACTYLY-MENTAL RETARDATION SYNDROME; BDMR
BJS	GDB:9955717	2q34-2q36	TORTI AND NERVE DEAFNESS
20 BMPR2	GDB:642243	2q33-2q33 2q33-2q34	PULMONARY HYPERTENSION, PRIMARY; PPH1 BONE MORPHOGENETIC RECEPTOR TYPE II; BMPR2
CHRNA1	GDB:120586	2q24-2q32	CHOLINERGIC RECEPTOR, NICOTINIC, ALPHA POLYPEPTIDE 1; CHRNA1
25 CMCWTD	GDB:11498919	2p22.3-2p21	FAMILIAL CHRONIC MUCOCUTANEOUS, DOMINANT TYPE
CNGA3	GDB:434398	2q11.2-2q11.2	COLORBLINDNESS, TOTAL CYCLIC NUCLEOTIDE GATED CHANNEL, OLFACTORY, 3; CNG3
30 COL3A1	GDB:118729	2q31-2q32.3 2q32.2-2q32.2	COLLAGEN, TYPE III; COL3A1 EHRLERS-DANLOS SYNDROME, TYPE IV, AUTOSOMAL DOMINANT
COL4A3	GDB:128351	2q36-2q37	COLLAGEN, TYPE IV, ALPHA-3 CHAIN; COL4A3
35 COL4A4	GDB:132673	2q35-2q37	COLLAGEN, TYPE IV, ALPHA-4 CHAIN; COL4A4

Gene	GDB Accession ID	Location	OMIM Link
COL6A3	GDB:119066	2q37.3-2q37.3	COLLAGEN, TYPE VI, ALPHA-3 CHAIN; COL6A3 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
CPS1	GDB:119799	2q33-2q36 2q34-2q35 2q35-2q35	HYPERAMMONEMIA DUE TO CARBAMOYLPHOSPHATE SYNTHETASE I DEFICIENCY
CRYGA	GDB:119076	2q33-2q35	CRYSTALLIN, GAMMA A; CRYGA
CRYGEP1	GDB:119808	2q33-2q35	CRYSTALLIN, GAMMA A; CRYGA
CYP1B1	GDB:353515	2p21-2p21 2p22-2p21 2pter-2qter	GLAUCOMA 3, PRIMARY INFANTILE, A; GLC3A CYTOCHROME P450, SUBFAMILY I (DIOXIN-INDUCIBLE), POLYPEPTIDE 1; CYP1B1
CYP27A1	GDB:128129	2q33-2qter	CEREBROTENDINOUS XANTHOMATOSIS
DBI	GDB:119837	2q12-2q21	DIAZEPAM BINDING INHIBITOR; DBI
DES	GDB:119841	2q35-2q35	DESMIN; DES
DYSF	GDB:340831	2p-2p 2p13-2p13 2pter-2p12	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2B; LGMD2B MUSCULAR DYSTROPHY, LATE-ONSET DISTAL
EDAR	GDB:9837372	2q11-2q13	DYSPLASIA, HYPOHIDROTIC ECTODERMAL DYSPLASIA, ANHIDROTIC
EFEMP1	GDB:1220111	2p16-2p16	DOYNE HONEYCOMB DEGENERATION OF RETINA FIBRILLIN-LIKE; FBNL
EIF2AK3	GDB:9956743	2p12-2p12	EPIPHYSEAL DYSPLASIA, MULTIPLE, WITH EARLY-ONSET DIABETES MELLITUS

Gene	GDB Accession ID	Location	OMIM Link
ERCC3	GDB:119881	2q21-2q21	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 3; ERCC3
FSHR	GDB:127510	2p21-2p16	FOLLICLE-STIMULATING HORMONE RECEPTOR; FSHR GONADAL DYSGENESIS, XX TYPE
GAD1	GDB:119244	2q31-2q31	PYRIDOXINE DEPENDENCY WITH SEIZURES
GINGF	GDB:9848875	2p21-2p21	GINGIVAL SON OF SEVENLESS (DROSOPHILA) HOMOLOG 1; SOS1
GLC1B	GDB:1297553	2q1-2q13	GLAUCOMA 1, OPEN ANGLE, B; GLC1B
GPD2	GDB:354558	2q24.1-2q24.1	GLYCEROL-3-PHOSPHATE DEHYDROGENASE-2; GPD2
GYPC	GDB:120027	2q14-2q21	BLOOD GROUP--GERBICH; Ge
HADHA	GDB:434026	2p23-2p23	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
HADHB	GDB:344953	2p23-2p23	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
HOXD13	GDB:127225	2q31-2q31	HOMEBOX-D13; HOXD13 SYNDACTYLY, TYPE II
HPE2	GDB:136066	2p21-2p21	MIDLINE CLEFT SYNDROME
IGKC	GDB:120088	2p12-2p12 2p11.2-2p11.2	IMMUNOGLOBULIN KAPPA CONSTANT REGION; IGKC
IHH	GDB:511203	2q33-2q35 2q35-2q35 2pter-2qter	BRACHYDACTYLY, TYPE A1; BDA1 INDIAN HEDGEHOG, DROSOPHILA, HOMOLOG OF; IHH
IRS1	GDB:133974	2q36-2q36	INSULIN RECEPTOR SUBSTRATE 1; IRS1

Gene	GDB Accession ID	Location	OMIM Link
ITGA6	GDB:128027	2pter-2qter	INTEGRIN, ALPHA-6; ITGA6
KHK	GDB:391903	2p23.3-2p23.2	FRUCTOSURIA
KYNU	GDB:9957925	2q22.2-2q23.3	
LCT	GDB:120140	2q21-2q21	DISACCHARIDE INTOLERANCE II
LHCGR	GDB:125260	2p21-2p21	LUTEINIZING HORMONE/CHORIOGONADOTROPIN RECEPTOR; LHCGR
LSFC	GDB:9956219	2-2 2p16-2p16	CYTOCHROME c OXIDASE DEFICIENCY, FRENCH-CANADIAN TYPE
MSH2	GDB:203983	2p16-2p16 2p22-2p21	COLON CANCER, FAMILIAL, NONPOLYPOSIS TYPE 1; FCC1
MSH6	GDB:632803	2p16-2p16	G/T MISMATCH-BINDING PROTEIN; GTBP
NEB	GDB:120224	2q24.1-2q24.2	NEBULIN; NEB NEMALINE MYOPATHY 2, AUTOSOMAL RECESSIVE; NEM2
NMTC	GDB:11498336	2q21-2q21	THYROID CARCINOMA, PAPILLARY
NPHP1	GDB:128050	2q13-2q13	NEPHRONOPHTHISIS, FAMILIAL JUVENILE 1; NPHP1
PAFAH1P1	GDB:435099	2p11.2-2p11.2	PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
PAX3	GDB:120495	2q36-2q36 2q35-2q35	KLEIN-WAARDENBURG SYNDROME WAARDENBURG SYNDROME; WS1
PAX8	GDB:136447	2q12-2q14	PAIRED BOX HOMEOTIC GENE 8; PAX8
PMS1	GDB:386403	2q31-2q33	POSTMEIOTIC SEGREGATION INCREASED (S. CEREVISIAE)-1; PMS1
PNKD	GDB:5583973	2q33-2q35	CHOREOATHETOSIS, FAMILIAL PAROXYSMAL; FPD1

Gene	GDB Accession ID	Location	OMIM Link
PPH1	GDB:1381541	2q31-2q32 2q33-2q33	PULMONARY HYPERTENSION, PRIMARY; PPH1
PROC	GDB:120317	2q13-2q21 2q13-2q14	PROTEIN C DEFICIENCY, CONGENITAL THROMBOTIC DISEASE DUE TO
REG1A	GDB:132455	2p12-2p12	REGENERATING ISLET-DERIVED 1-ALPHA; REG1A
SAG	GDB:120365	2q37.1-2q37.1	S-ANTIGEN; SAG
SFTPB	GDB:120374	2p12-2p11.2	SURFACTANT-ASSOCIATED PROTEIN, PULMONARY-3; SFTP3
SLC11A1	GDB:371444	2q35-2q35	CIRRHOSIS, PRIMARY; PBC NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 1; NRAMP1
SLC3A1	GDB:202968	2p16.3-2p16.3 2p21-2p21	SOLUTE CARRIER FAMILY 3, MEMBER 1; SLC3A1 CYSTINURIA; CSNU
SOS1	GDB:230004	2p22-2p21	GINGIVAL SON OF SEVENLESS (DROSOPHILA) HOMOLOG 1; SOS1
SPG4	GDB:230127	2p24-2p21	SPASTIC PARAPLEGIA-4, AUTOSOMAL DOMINANT; SPG4
SRD5A2	GDB:127343	2p23-2p23	PSEUDOVAGINAL PERINEOSCROTAL HYPOSPADIAS; PPSH
TCL4	GDB:136378	2q34-2q34	T-CELL LEUKEMIA/LYMPHOMA-4; TCL4
TGFA	GDB:120435	2p13-2p13	TRANSFORMING GROWTH FACTOR, ALPHA; TGFA
TMD	GDB:9837196	2q31-2q31	TIBIAL MUSCULAR DYSTROPHY, TARDIVE

	Gene	GDB Accession ID	Location	OMIM Link
5	TPO	GDB:120446	2p25-2p25 2p25-2p24	THYROID HORMONOGENESIS, GENETIC DEFECT IN, IIA
10	UGT1	GDB:120007	2q37-2q37	UDP GLUCURONOSYLTRANSFERASE 1 FAMILY, A1; UGT1A1
15	UV24	GDB:9955737	2pter-2qter	UV-DAMAGE, EXCISION REPAIR OF, UV-24
20	WSS	GDB:9955707	2q32-2q32	WRINKLY SKIN SYNDROME; WSS
25	XDH	GDB:266386	2p23-2p22	XANTHINURIA
30	ZAP70	GDB:433738	2q11-2q13 2q12-2q12	SYK-RELATED TYROSINE KINASE; SRK
35	ZFHX1B	GDB:9958310	2q22-2q22	DISEASE, MICROCEPHALY, AND IRIS COLOBOMA

Table 4: Genes, Locations and Genetic Disorders on Chromosome 3

	Gene	GDB Accession ID	Location	OMIM Link
20	ACAA1	GDB:119643	3p23-3p22	PEROXISOMAL 3-OXOACYL-COENZYME A THIOLASE DEFICIENCY
25	AGTR1	GDB:132359	3q21-3q25	ANGIOTENSIN II RECEPTOR, VASCULAR TYPE 1; AT2R1
30	AHSG	GDB:118985	3q27-3q27	ALPHA-2-HS-GLYCOPROTEIN; AHSG
35	AMT	GDB:132138	3p21.3-3p21.2 3p21.2-3p21.1	HYPERGLYCINEMIA, ISOLATED NONKETOTIC, TYPE II; NKH2
	ARP	GDB:9959049	3p21.1-3p21.1	ARGININE-RICH PROTEIN
	BBS3	GDB:376501	3p-3p 3p12.3-3q11.1	BARDET-BIEDL SYNDROME, TYPE 3; BBS3
	BCHE	GDB:120558	3q26.1-3q26.2	BUTYRYLCHOLINESTERASE; BCHE
	BCPM	GDB:433809	3q21-3q21	BENIGN CHRONIC PEMPHIGUS; BCPM
	BTD	GDB:309078	3p25-3p25	BIOTINIDASE; BTD

Gene	GDB Accession ID	Location	OMIM Link
CASR	GDB:134196	3q21-3q24	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL; HHC1
5 CCR2	GDB:337364	3p21-3p21	CHEMOKINE (C-C) RECEPTOR 2; CMKBR2
CCR5	GDB:1230510	3p21-3p21	CHEMOKINE (C-C) RECEPTOR 5; CMKBR5
10 CDL1	GDB:136344	3q26.3-3q26.3	DE LANGE SYNDROME; CDL
CMT2B	GDB:604021	3q13-3q22	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, B; CMT2B
15 COL7A1	GDB:128750	3p21-3p21 3p21.3-3p21.3	COLLAGEN, TYPE VII, ALPHA-1; COL7A1
CP	GDB:119069	3q23-3q25 3q21-3q24	CERULOPLASMIN; CP
CRV	GDB:11498333	3p21.3-3p21.1	VASCULOPATHY, RETINAL, WITH CEREBRAL LEUKODYSTROPHY
20 CTNNB1	GDB:141922	3p22-3p22 3p21.3-3p21.3	CATENIN, BETA 1; CTNNB1
DEM	GDB:681157	3p12-3q11	DEMENTIA, FAMILIAL NONSPECIFIC; DEM
ETM1	GDB:9732523	3q13-3q13	TREMOR, HEREDITARY ESSENTIAL 1; ETM1
25 FANCD2	GDB:698345	3p25.3-3p25.3 3pter-3p24.2	FANCONI PANCYTOPENIA, COMPLEMENTATION GROUP D
FIH	GDB:9955790	3q13-3q13	HYPOPARATHYROIDISM, FAMILIAL ISOLATED; FIH
30 FOXL2	GDB:129025	3q23-3q23 3q22-3q23	BLEPHAROPHIMOSIS, EPICANTHUS INVERSUS, AND PTOSIS; BPES
GBE1	GDB:138442	3p12-3p12	GLYCOGEN STORAGE DISEASE IV
GLB1	GDB:119987	3p22-3p21.33 3p21.33-3p21.33	GANGLIOSIDOSIS, GENERALIZED GM1, TYPE I

Gene	GDB Accession ID	Location	OMIM Link
GLC1C	GDB:3801941	3q21-3q24	GLAUCOMA 1, OPEN ANGLE, C; GLC1C
5 GNAI2	GDB:120516	3p21.3-3p21.2	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-INHIBITING, POLYPEPTIDE-2;
10 GNAT1	GDB:119277	3p21.3-3p21.2	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-TRANSDUCING, POLYPEPTIDE
15 GP9	GDB:126370	3pter-3qter	PLATELET GLYCOPROTEIN IX; GP9
15 ITIH1	GDB:119282	3q11-3q12 3p21.3-3p21.3	GLUTATHIONE PEROXIDASE; GPX1
20 HGD	GDB:203935	3q21-3q23	ALKAPTONURIA; AKU
20 HRG	GDB:120055	3q27-3q27	HISTIDINE-RICH GLYCOPROTEIN; HRG; HRGP
20 ITIH1	GDB:120107	3p21.2-3p21.1	INTER-ALPHA-TRYPSIN INHIBITOR, HEAVY CHAIN-1; ITIH1; LATIH; ITIH
25 KNG	GDB:125256	3q27-3q27	FLAUJEAC FACTOR DEFICIENCY
25 LPP	GDB:1391795	3q27-3q28	LIM DOMAIN-CONTAINING PREFERRED TRANSLOCATION PARTNER IN LIPOMA; LPP
30 LRS1	GDB:682448	3p21.1-3p14.1	LARSEN SYNDROME, AUTOSOMAL DOMINANT; LRS1
30 MCCC1	GDB:135989	3q27-3q27 3q25-3q27	BETA-METHYLCROTONYLGLYCINURIA I
30 MDS1	GDB:250411	3q26-3q26	MYELODYSPLASIA SYNDROME 1; MDS1
35 MHS4	GDB:574245	3q13.1-3q13.1	HYPERTHERMIA SUSCEPTIBILITY-4; MHS4
35 MITF	GDB:214776	3p14.1-3p12	MICROPHTHALMIA-ASSOCIATED TRANSCRIPTION FACTOR; MITF WAARDENBURG SYNDROME, TYPE II; WS2

	Gene	GDB Accession ID	Location	OMIM Link
	MLH1	GDB:249617	3p23-3p22 3p21.3-3p21.3	COLON CANCER, FAMILIAL, NONPOLYPOSIS TYPE 2; FCC2
5	MYL3	GDB:120218	3p21.3-3p21.2	MYOSIN, LIGHT CHAIN, ALKALI, VENTRICULAR AND SKELETAL SLOW; MYL3
	MYMY	GDB:11500610	3p26-3p24.2	DISEASE
	OPA1	GDB:118848	3q28-3q29	OPTIC ATROPHY 1; OPA1
10	PBXP1	GDB:125352	3q22-3q23	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-1; PBX1
	PCCB	GDB:119474	3q21-3q22	GLYCINEMIA, KETOTIC, II
15	POU1F1	GDB:129070	3p11-3p11	POU DOMAIN, CLASS 1, TRANSCRIPTION FACTOR 1; POU1F1
	PPARG	GDB:1223810	3p25-3p25	CANCER OF COLON PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR, GAMMA; PPARG
20	PROS1	GDB:120721	3p11-3q11 3p11.1-3q11.2	PROTEIN S, ALPHA; PROS1
	PTHR1	GDB:138128	3p22-3p21.1	METAPHYSEAL CHONDRODYSPLASIA, MURK JANSEN TYPE PARATHYROID HORMONE RECEPTOR 1; PTHR1
25	RCA1	GDB:230233	3p14.2-3p14.2	RENAL CARCINOMA, FAMILIAL, ASSOCIATED 1; RCA1
	RHO	GDB:120347	3q21.3-3q24	RHODOPSIN; RHO
	SCA7	GDB:454471	3p21.1-3p12	SPINOCEBELLAR ATAXIA 7; SCA7
30	SCLC1	GDB:9955750	3p23-3p21	SMALL-CELL CANCER OF THE LUNG; SCCL
	SCN5A	GDB:132152	3p21-3p21	SODIUM CHANNEL, VOLTAGE-GATED, TYPE V, ALPHA POLYPEPTIDE; SCN5A
	SI	GDB:120377	3q25.2-3q26.2	DISACCHARIDE INTOLERANCE I

	Gene	GDB Accession ID	Location	OMIM Link
	SLC25A20	GDB:6503297	3p21.31-3p21.31	CARNITINE-ACYLCARNITINE TRANSLOCASE; CACT
5	SLC2A2	GDB:119995	3q26.2-3q27 3q26.1-3q26.3	SOLUTE CARRIER FAMILY 2, MEMBER 2; SLC2A2 FANCONI-BICKEL SYNDROME; FBS
	TF	GDB:120432	3q21-3q21	TRANSFERRIN; TF
10	TGFB2	GDB:224909	3p22-3p22 3pter-3p24.2	TRANSFORMING GROWTH FACTOR-BETA RECEPTOR, TYPE II; TGFB2
	THPO	GDB:374007	3q26.3-3q27	THROMBOPOIETIN; THPO
	THRB	GDB:120731	3p24.1-3p22 3p24.3-3p24.3	THYROID HORMONE RECEPTOR, BETA; THRB
15	TKT	GDB:132402	3p14.3-3p14.3	WERNICKE-KORSAKOFF SYNDROME
	TM4SF1	GDB:250815	3q21-3q25	TUMOR-ASSOCIATED ANTIGEN L6; TAAL6
20	TRH	GDB:128072	3pter-3qter	THYROTROPIN-RELEASING HORMONE DEFICIENCY
	UMPS	GDB:120482	3q13-3q13	OROTICACIDURIA I
	UQCRC1	GDB:141850	3p21.3-3p21.2 3p21.3-3p21.3	UBIQUINOL-CYTOCHROME c REDUCTASE CORE PROTEIN I; UQCRC1
25	USH3A	GDB:392645	3q21-3q25	USHER SYNDROME, TYPE III; USH3
	VHL	GDB:120488	3p26-3p25	VON HIPPEL-LINDAU SYNDROME; VHL
30	WS2A	GDB:128053	3p14.2-3p13	MICROPHTHALMIA-ASSOCIATED TRANSCRIPTION FACTOR; MITF WAARDENBURG SYNDROME, TYPE II; WS2
	XPC	GDB:134769	3p25.1-3p25.1	XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP C; XPC
35	ZNF35	GDB:120507	3p21-3p21	ZINC FINGER PROTEIN-35; ZNF35

Table 5: Genes, Locations and Genetic Disorders on Chromosome 4

	Gene	GDB Accession ID	Location	OMIM Link
5	ADH1B	GDB:119651	4q21-4q23 4q22-4q22	ALCOHOL DEHYDROGENASE-2; ADH2
	ADH1C	GDB:119652	4q21-4q23 4q22-4q22	ALCOHOL DEHYDROGENASE-3; ADH3
10	AFP	GDB:119660	4q11-4q13	ALPHA-FETOPROTEIN; AFP
	AGA	GDB:118981	4q23-4q35 4q32-4q33	ASPARTYLGLUCOSAMINURIA ; AGU
15	AIH2	GDB:118751	4q11-4q13 4q13.3-4q21.2	AMELOGENESIS IMPERFECTA 2, HYPOPLASTIC LOCAL, AUTOSOMAL DOMINANT;
	ALB	GDB:118990	4q11-4q13	ALBUMIN; ALB
20	ASMD	GDB:119705	4q-4q 4q28-4q31	ANTERIOR SEGMENT OCULAR DYSGENESIS; ASOD
	BFHD	GDB:11498907	4q34.1-4q35	DYSPLASIA, BEUKES TYPE
25	CNGA1	GDB:127557	4p14-4q13	CYCLIC NUCLEOTIDE GATED CHANNEL, PHOTORECEPTOR, cGMP GATED, 1; CNCG1
	CRBM	GDB:9958132	4p16.3-4p16.3	CHERUBISM
30	DCK	GDB:126810	4q13.3-4q21.1	DEOXYCYTIDINE KINASE; DCK
	DFNA6	GDB:636175	4p16.3-4p16.3	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 6; DFNA6
35	DSPP	GDB:5560457	4pter-4qter 4q21.3-4q21.3	DENTIN PHOSPHOPROTEIN; DPP DENTINOGENESIS IMPERFECTA; DGI1
	DTDP2	GDB:9955810	4q-4q	DENTIN DYSPLASIA, TYPE II
40	ELONG	GDB:11498700	4q24-4q24	
	ENAM	GDB:9955259	4q21-4q21	AMELOGENESIS IMPERFECTA 2, HYPOPLASTIC LOCAL, AUTOSOMAL DOMINANT; AMELOGENESIS IMPERFECTA, HYPOPLASTIC TYPE
45	ETFDH	GDB:135992	4q32-4q35	GLUTARICACIDURIA IIC; GA IIC

Gene	GDB Accession ID	Location	OMIM Link
EVC	GDB:555573	4p16-4p16	ELLIS-VAN CREVELD SYNDROME; EVC
5 F11	GDB:119891	4q35-4q35	PTA DEFICIENCY
FABP2	GDB:119127	4q28-4q31	FATTY ACID BINDING PROTEIN 2, INTESTINAL; FABP2
10 FGA	GDB:119129	4q28-4q28	AMYLOIDOSIS, FAMILIAL; VISCERAL FIBRINOGEN, A ALPHA POLYPEPTIDE; FGA
FGB	GDB:119130	4q28-4q28	FIBRINOGEN, B BETA POLYPEPTIDE; FGB
15 FGFR3	GDB:127526	4p16.3-4p16.3	ACHONDROPLASIA; ACH BLADDER CANCER; FIBROBLAST GROWTH FACTOR RECEPTOR-3; FGFR3
FGG	GDB:119132	4q28-4q28	FIBRINOGEN, G GAMMA POLYPEPTIDE; FGG
20 FSHMD1A	GDB:119914	4q35-4q35	FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY 1A; FSHMD1A
GC	GDB:119263	4q12-4q13 4q12-4q12	GROUP-SPECIFIC COMPONENT; GC
GNPTA	GDB:119280	4q21-4q23	MUCOLIPIDOSIS II; ML2; ML II
25 GNRHR	GDB:136456	4q13-4q13 4q21.2-4q21.2	GONADOTROPIN-RELEASING HORMONE RECEPTOR; GNRHR
GYPA	GDB:118890	4q28-4q31 4q28.2-4q31.1	BLOOD GROUP--MN LOCUS; MN
HCA	GDB:9954675	4q33-4qter	HYPERCALCIURIA, FAMILIAL IDIOPATHIC
30 HCL2	GDB:119305	4q28-4q31 4q-4q	HAIR COLOR-2; HCL2
HD	GDB:119307	4p16.3-4p16.3	HUNTINGTON DISEASE; HD
HTN3	GDB:125601	4q12-4q21	HISTATIN-3; HTN3
HVBS6	GDB:120687	4q32-4q32	HEPATOCELLULAR CARCINOMA-2; HCC2
35 IDUA	GDB:119327	4p16.3-4p16.3	MUCOPOLYSACCHARIDOSIS TYPE I; MPS I

Gene	GDB Accession ID	Location	OMIM Link
IF	GDB:120077	4q24-4q25 4q25-4q25	COMPLEMENT COMPONENT-3 INACTIVATOR, DEFICIENCY OF
5 JPD	GDB:120113	4pter-4qter 4q12-4q13	PERIODONTITIS, JUVENILE; JPD
KIT	GDB:120117	4q12-4q12	V-KIT HARDY-ZUCKERMAN 4 FELINE SARCOMA VIRAL ONCOGENE HOMOLOG; KIT
10 KLKB1	GDB:127575	4q34-4q35 4q35-4q35	FLETCHER FACTOR DEFICIENCY
LQT4	GDB:682072	4q25-4q27	SYNDROME WITHOUT PSYCHOMOTOR RETARDATION
15 MANBA	GDB:125261	4q21-4q25	MANNOSIDOSIS, BETA; MANB1
MLLT2	GDB:136792	4q21-4q21	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 2; MLLT2
20 MSX1	GDB:120683	4p16.3-4p16.1 4p16.1-4p16.1	MSH, DROSOPHILA, HOME BOX, HOMOLOG OF, 1; MSX1
MTP	GDB:228961	4q24-4q24	MICROSOMAL TRIGLYCERIDE TRANSFER PROTEIN, 88 KD; MTP
25 NR3C2	GDB:120188	4q31-4q31 4q31.1-4q31.1	PSEUDOHYPOALDOSTERONIS M, TYPE I, AUTOSOMAL RECESSIVE; PHA1
PBT	GDB:120260	4q12-4q21	PIEBALD TRAIT; PBT
30 PDE6B	GDB:125915	4p16.3-4p16.3	NIGHTBLINDNESS, CONGENITAL STATIONARY; CSNB3 PHOSPHODIESTERASE 6B, cGMP-SPECIFIC, ROD, BETA; PDE6B
PEE1	GDB:7016765	4q31-4q34 4q25-4qter	1; PEE1

Gene	GDB Accession ID	Location	OMIM Link
PITX2	GDB:134770	4q25-4q27 4q25-4q26 4q25-4q25	IRIDOGONIODYSGENESIS, TYPE 2; IRID2 RIEGER SYNDROME, TYPE 1; RIEG1 RIEG BICOID-RELATED HOMEobox TRANSCRIPTION FACTOR 1; RIEG1 HOME BOX 2
5			
PKD2	GDB:118851	4q21-4q23	POLYCYSTIC KIDNEY DISEASE 2; PKD2
10			
QDPR	GDB:120331	4p15.3-4p15.3 4p15.31-4p15.31	PHENYLKETONURIA II
15			
SGCB	GDB:702072	4q12-4q12	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2E; LGMD2E
20			
SLC25A4	GDB:119680	4q35-4q35	ADENINE NUCLEOTIDE TRANSLOCATOR 1; ANT1 PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA; PEO
25			
SNCA	GDB:439047	4q21.3-4q22 4q21-4q21	SYNUCLEIN, ALPHA; SNCA PARKINSON DISEASE, FAMILIAL, TYPE 1; PARK1
30			
SOD3	GDB:125291	4p16.3-4q21	SUPEROXIDE DISMUTASE, EXTRACELLULAR; SOD3
STATH	GDB:120391	4q11-4q13	STATHERIN; STATH; STR
TAPVR1	GDB:392646	4p13-4q11	ANOMALOUS PULMONARY VENOUS RETURN; APVR
TYS	GDB:119624	4q-4q	SCLEROTYLOSIS; TYS
WBS2	GDB:132426	4q33-4q35.1	WILLIAMS-BEUREN SYNDROME; WBS
WFS1	GDB:434294	4p-4p 4p16-4p16	DIABETES MELLITUS AND INSIPIDUS WITH OPTIC ATROPHY AND DEAFNESS
WHCR	GDB:125355	4p16.3-4p16.3	WOLF-HIRSCHHORN SYNDROME; WHS

Table 6: Genes, Locations and Genetic Disorders on Chromosome 5

Gene	GDB Accession ID	OMIM Link
5 ADAMTS2	GDB:9957209	EHLERS-DANLOS SYNDROME, TYPE VII, AUTOSOMAL RECESSIVE
10 ADRB2	GDB:120541	BETA-2-ADRENERGIC RECEPTOR; ADRB2
15 AMCN	GDB:9836823	ARTHROGRYPOSIS MULTIPLEX CONGENITA, NEUROGENIC TYPE
20 AP3B1	GDB:9955590	HERMANSKY-PUDLAK SYNDROME; HPS
25 APC	GDB:119682	ADENOMATOUS POLYPOSIS OF THE COLON; APC
30 ARSB	GDB:119008	MUCOPOLYSACCHARIDOSIS TYPE VI; MPS VI
B4GALT7	GDB:9957653	SYNDROME, PROGEROID FORM
BHR1	GDB:9956078	ASTHMA
C6	GDB:119045	COMPLEMENT COMPONENT-6, DEFICIENCY OF
C7	GDB:119046	COMPLEMENT COMPONENT-7, DEFICIENCY OF
CCAL2	GDB:5584265	CHONDROCALCINOSIS, FAMILIAL ARTICULAR
CKN1	GDB:128586	COCKAYNE SYNDROME, TYPE I; CKN1
CMDJ	GDB:9595425	CRANIOMETAPHYSEAL DYSPLASIA, JACKSON TYPE; CMDJ
CRHBP	GDB:127438	CORTICOTROPIN RELEASING HORMONE-BINDING PROTEIN; CRHBP
CSF1R	GDB:120600	COLONY-STIMULATING FACTOR-1 RECEPTOR; CSF1R
DHFR	GDB:119845	DIHYDROFOLATE REDUCTASE; DHFR

Gene	GDB Accession ID	OMIM Link
DIAPH1	GDB:9835482	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 1; DFNA1 DIAPHANOUS, DROSOPHILA, HOMOLOG OF, 1
DTR	GDB:119853	DIPHTHERIA TOXIN SENSITIVITY; DTS
EOS	GDB:9956083	EOSINOPHILIA, FAMILIAL
10 ERVR	GDB:9835857	HYALOIDEORETINAL DEGENERATION OF WAGNER
F12	GDB:119892	HAGEMAN FACTOR DEFICIENCY
15 FBN2	GDB:128122	CONTRACTURAL ARACHNODACTYLY, CONGENITAL; CCA
GDNF	GDB:450609	GLIAL CELL LINE-DERIVED NEUROTROPHIC FACTOR; GDNF
GHR	GDB:119984	GROWTH HORMONE RECEPTOR; GHR
20 GLRA1	GDB:118801	GLYCINE RECEPTOR, ALPHA-1 SUBUNIT; GLRA1 KOK DISEASE
GM2A	GDB:120000	TAY-SACHS DISEASE, AB VARIANT
HEXB	GDB:119308	SANDHOFF DISEASE
25 HSD17B4	GDB:385059	17-@BETA-HYDROXYSTEROID DEHYDROGENASE IV; HSD17B4
ITGA2	GDB:128031	INTEGRIN, ALPHA-2; ITGA2
KFS	GDB:9958987	VERTEBRAL FUSION
LGMD1A	GDB:118832	MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 1A; LGMD1A
30 LOX	GDB:119367	LYSYL OXIDASE; LOX
LTC4S	GDB:384080	LEUKOTRIENE C4 SYNTHASE; LTC4S
35 MAN2A1	GDB:136413	MANNOSIDASE, ALPHA, II; MANA2 DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE II

	<b>Gene</b>	<b>GDB Accession ID</b>	<b>OMIM Link</b>
	MCC	GDB:128163	MUTATED IN COLORECTAL CANCERS; MCC
5	MCCC2	GDB:135990	II
	MSH3	GDB:641986	MutS, E. COLI, HOMOLOG OF, 3; MSH3
	MSX2	GDB:138766	MSH (DROSOPHILA) HOMEO BOX HOMOLOG 2; MSX2 PARIELTAL FORAMINA, SYMMETRIC; PFM
10	NR3C1	GDB:120017	GLUCOCORTICOID RECEPTOR; GRL
	PCSK1	GDB:128033	PROPROTEIN CONVERTASE SUBTILISIN/KEXIN TYPE 1; PCSK1
15	PDE6A	GDB:120265	PHOSPHODIESTERASE 6A, cGMP-SPECIFIC, ROD, ALPHA; PDE6A
	PFBI	GDB:9956096	INTENSITY OF INFECTION IN
	RASA1	GDB:120339	RAS p21 PROTEIN ACTIVATOR 1; RASA1
20	SCZD1	GDB:120370	DISORDER-1; SCZD1
	SDHA	GDB:378037	SUCCINATE DEHYDROGENASE COMPLEX, SUBUNIT A, FLAVOPROTEIN; SDHA
	SGCD	GDB:5886421	SARCOGLYCAN, DELTA; SGCD
25	SLC22A5	GDB:9863277	CARNITINE DEFICIENCY, SYSTEMIC, DUE TO DEFECT IN RENAL REABSORPTION
	SLC26A2	GDB:125421	DIASTROPHIC DYSPLASIA; DTD EPIPHYSEAL DYSPLASIA, MULTIPLE; MED NEONATAL OSSEOUS DYSPLASIA I; ACHONDROGENESIS, TYPE IB; ACG1B
30	SLC6A3	GDB:132445	SOLUTE CARRIER FAMILY 6, MEMBER 3; SLC6A3 DEFICIT-HYPERACTIVITY DISORDER; ADHD
	SM1	GDB:9834488	SCHISTOSOMA MANSONI SUSCEPTIBILITY/RESISTANCE

Gene	GDB Accession ID	OMIM Link
SMA@	GDB:120378	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 1, TELOMERIC; SMN1
5 SMN1	GDB:5215173	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 1, TELOMERIC; SMN1
10 SMN2	GDB:5215175	SPINAL MUSCULAR ATROPHY I; SMA I SURVIVAL OF MOTOR NEURON 2, CENTROMERIC; SMN2
SPINK5	GDB:9956114	NETHERTON DISEASE
15 TCOF1	GDB:127390	TREACHER COLLINS-FRANCESCHETTI SYNDROME 1; TCOF1
TGFBI	GDB:597601	CORNEAL DYSTROPHY, GRANULAR TYPE CORNEAL DYSTROPHY, LATTICE TYPE I; CDL1 TRANSFORMING GROWTH FACTOR, BETA-INDUCED, 68 KD; TGFBI

Table 7: Genes, Locations and Genetic Disorders on Chromosome 6

Gene	GDB Accession ID	OMIM Link
20 ALDH5A1	GDB:454767	SUCCINIC SEMIALDEHYDE DEHYDROGENASE, NAD(+)-DEPENDENT; SSADH
ARG1	GDB:119006	ARGININEMIA
25 AS	GDB:135697	ANKYLOSING SPONDYLITIS; AS
ASSP2	GDB:119017	CITRULLINEMIA
BCKDHB	GDB:118759	MAPLE SYRUP URINE DISEASE, TYPE IB
30 BF	GDB:119726	GLYCINE-RICH BETA-GLYCOPROTEIN; GBG
C2	GDB:119731	COMPLEMENT COMPONENT-2, DEFICIENCY OF
C4A	GDB:119732	COMPLEMENT COMPONENT 4A; C4A
35 CDKN1A	GDB:266550	CYCLIN-DEPENDENT KINASE INHIBITOR 1A; CDKN1A

	<b>Gene</b>	<b>GDB Accession ID</b>	<b>OMIM Link</b>
	COL10A1	GDB:128635	COLLAGEN, TYPE X, ALPHA 1; COL10A1
5	COL11A2	GDB:119788	COLLAGEN, TYPE XI, ALPHA-2; COL11A2 STICKLER SYNDROME, TYPE II; STL2 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 13; DFNA13
10	CYP21A2	GDB:120605	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 21-HYDROXYLASE DEFICIENCY
	DYX2	GDB:437584	DYSLEXIA, SPECIFIC, 2; DYX2
	EJM1	GDB:119864	MYOCLONIC EPILEPSY, JUVENILE; EJM1
15	ELOVL4	GDB:11499609	STARGARDT DISEASE 3; STGD3
	EPM2A	GDB:3763331	EPILEPSY, PROGRESSIVE MYOCLONIC 2; EPM2
	ESR1	GDB:119120	ESTROGEN RECEPTOR; ESR
20	EYA4	GDB:700062	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 10; DFNA10
	F13A1	GDB:120614	FACTOR XIII, A1 SUBUNIT; F13A1
	FANCE	GDB:1220236	FANCONI ANEMIA, COMPLEMENTATION GROUP E; FACE
25	GCLC	GDB:132915	GAMMA-GLUTAMYL CYSTEINE SYNTHETASE DEFICIENCY, HEMOLYTIC ANEMIA DUE
	GJA1	GDB:125196	GAP JUNCTION PROTEIN, ALPHA-1, 43 KD; GJA1
	GLYS1	GDB:136421	GLYCOSURIA, RENAL
30	GMPR	GDB:127058	GUANINE MONOPHOSPHATE REDUCTASE
	GSE	GDB:9956235	DISEASE; CD
	HCR	GDB:9993306	PSORIASIS, SUSCEPTIBILITY TO HFE GDB:119309 HEMOCHROMATOSIS; HFE
35			

Gene	GDB Accession ID	OMIM Link
5 HLA-A	GDB:119310	HLA-A HISTOCOMPATIBILITY TYPE; HLA-A HLA-DPB1GDB:120636 LA-DP HISTOCOMPATIBILITY TYPE, BETA-1 SUBUNIT
10 HLA-DRA	GDB:120641	HLA-DR HISTOCOMPATIBILITY TYPE; HLA-DRA
15 HPFH	GDB:9849006	HETEROCELLULAR HEREDITARY PERSISTENCE OF FETAL HEMOGLOBIN
20 ICS1	GDB:136433	IMMOTILE CILIA SYNDROME-1; ICS1
25 IDDM1	GDB:9953173	DIABETES MELLITUS, JUVENILE-ONSET INSULIN-DEPENDENT; IDDM
30 IFNGR1	GDB:120688	INTERFERON, GAMMA, RECEPTOR-1; IFNGR1
35 IGAD1	GDB:6929077	SELECTIVE DEFICIENCY OF
LAMA2	GDB:120083	INSULIN-LIKE GROWTH FACTOR 2 RECEPTOR; IGF2R
ISCW	GDB:9956158	SUPPRESSION; IS
LAMA2	GDB:132362	LAMININ, ALPHA 2; LAMA2
LAP	GDB:9958992	LARYNGEAL ADDUCTOR PARALYSIS; LAP
LCA5	GDB:11498764	AMAUROSIS CONGENITA OF LEBER I
LPA	GDB:120699	APOLIPOPROTEIN(a); LPA
MCDR1	GDB:131406	MACULAR DYSTROPHY, RETINAL, 1, NORTH CAROLINA TYPE; MCDR1
MOCS1	GDB:9862235	MOLYBDENUM COFACTOR DEFICIENCY
MUT	GDB:120204	METHYLMALONICACIDURIA DUE TO METHYLMALONIC CoA MUTASE DEFICIENCY
MYB	GDB:119441	V-MYB AVIAN MYELOBLASTOSIS VIRAL ONCOGENE HOMOLOG; MYB
NEU1	GDB:120230	NEURAMINIDASE DEFICIENCY

Gene	GDB Accession ID	OMIM Link
NKS1	GDB:128100	SUSCEPTIBILITY TO LYSIS BY ALLOREACTIVE NATURAL KILLER CELLS; EC1
NYS2	GDB:9848763	NYSTAGMUS, CONGENITAL
OA3	GDB:136429	ALBINISM, OCULAR, AUTOSOMAL RECESSIVE; OAR
ODDD	GDB:6392584	OCULODENTODIGITAL DYSPLASIA; ODDD
OFC1	GDB:120247	OROFACIAL CLEFT 1; OFC1
PARK2	GDB:6802742	PARKINSONISM, JUVENILE
PBCA	GDB:9956321	BETA CELL AGENESIS WITH NEONATAL DIABETES MELLITUS
PBCRA1	GDB:3763333	CHORIORETINAL ATROPHY, PROGRESSIVE BIFOCAL; CRAPB
PDB1	GDB:136349	DISEASE OF BONE; PDB
PEX3	GDB:9955507	ZELLWEGER SYNDROME; ZS
PEX6	GDB:5592414	ZELLWEGER SYNDROME; ZS PEROXIN-6; PEX6
PEX7	GDB:6155803	RHIZOMELIC CHONDRODYSPLASIA PUNCTATA; RCDP PEROXIN-7; PEX7
PKHD1	GDB:433910	POLYCYSTIC KIDNEY AND HEPATIC DISEASE-1; PKHD1
PLA2G7	GDB:9958829	PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, SUBUNIT
PLG	GDB:119498	PLASMINOGEN; PLG
POLH	GDB:6963323	PIGMENTOSUM WITH NORMAL DNA REPAIR RATES
PPAC	GDB:9956248	ARTHROPATHY, PROGRESSIVE PSEUDORHEUMATOID, OF CHILDHOOD
PSORS1	GDB:6381310	PSORIASIS, SUSCEPTIBILITY TO
PUJO	GDB:9956231	MULTICYSTIC RENAL DYSPLASIA, BILATERAL; MRD
RCD1	GDB:333929	RETINAL CONE DEGENERATION

Gene	GDB Accession ID	OMIM Link
RDS	GDB:118863	RETINAL DEGENERATION, SLOW; RDS
5 RHAG	GDB:136011	RHESUS BLOOD GROUP-ASSOCIATED GLYCOPROTEIN; RHAG RH-NULL, REGULATOR TYPE; RHN
RP14	GDB:433713	RETINITIS PIGMENTOSA-14; RP14 TUBBY-LIKE PROTEIN 1; TULP1
10 RUNX2	GDB:392082	CLEIDOCRANIAL DYSPLASIA; CCD CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 1; CBFA1
RWS	GDB:9956195	SENSITIVITY
SCA1	GDB:119588	SPINOCEREBELLAR ATAXIA 1; SCA1
15 SCZD3	GDB:635974	DISORDER-3; SCZD3
SIASD	GDB:433552	SIALIC ACID STORAGE DISEASE; SIASD
SOD2	GDB:119597	SUPEROXIDE DISMUTASE 2, MITOCHONDRIAL; SOD2
20 ST8	GDB:6118456	OVARIAN TUMOR
TAP1	GDB:132668	TRANSPORTER 1, ABC; TAP1
TAP2	GDB:132669	TRANSPORTER 2, ABC; TAP2
25 TFAP2B	GDB:681506	DUCTUS ARTERIOSUS; PDA TRANSCRIPTION FACTOR AP-2 BETA; TFAP2B
TNDM	GDB:9956265	DIABETES-MELLITUS, TRANSIENT NEONATAL
TNF	GDB:120441	TUMOR NECROSIS FACTOR; TNF
30 TPBG	GDB:125568	TROPHOBlast GLYCOPROTEIN; TPBG; M6P1
TPMT	GDB:209025	THIOPURINE S-METHYLTRANSFERASE; TPMT
TULP1	GDB:6199353	TUBBY-LIKE PROTEIN 1; TULP1
35 WISP3	GDB:9957361	ARTHROPATHY, PROGRESSIVE PSEUDORHEUMATOID, OF CHILDHOOD

Table 8: Genes, Locations and Genetic Disorders on Chromosome 7

	Gene	GDB Accession ID	OMIM Link
5	AASS	GDB:11502144	HYPERLYSINEMIA
	ABCB1	GDB:120712	P-GLYCOPROTEIN-1; PGY1
	ABCB4	GDB:120713	P-GLYCOPROTEIN-3; PGY3
10	ACHE	GDB:118746	ACETYLCHOLINESTERASE BLOOD GROUP--Yt SYSTEM; YT
	AQP1	GDB:129082	AQUAPORIN-1; AQP1 BLOOD GROUP--COLTON; CO
	ASL	GDB:119703	ARGININOSUCCINICACIDURIA
	ASNS	GDB:119706	ASPARAGINE SYNTHETASE; ASNS; AS
15	AUTS1	GDB:9864226	DISORDER
	BPGM	GDB:119039	DIPHOSPHOGLYCERATE MUTASE DEFICIENCY OF ERYTHROCYTE
	C7orf2	GDB:10794644	ACHEIROPODY
20	CACNA2D1	GDB:132010	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, L TYPE, ALPHA-2/DELTA SUBUNIT; MALIGNANT HYPERTHERMIA SUSCEPTIBILITY-3
	CCM1	GDB:580824	CEREBRAL CAVERNOUS MALFORMATIONS 1; CCM1
25	CD36	GDB:138800	CD36 ANTIGEN; CD36
	CFTR	GDB:120584	CYSTIC FIBROSIS; CF DEFERENS, CONGENITAL BILATERAL APLASIA OF; CBAVD; CAVD
	CHORDOMA	GDB:11498328	
30	CLCN1	GDB:134688	CHLORIDE CHANNEL 1, SKELETAL MUSCLE; CLCN1
	CMH6	GDB:9956392	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, WITH WOLFF-PARKINSON-WHITE
35	CMT2D	GDB:9953232	CHARCOT-MARIE-TOOTH DISEASE, NEURONAL TYPE, D

Gene	GDB Accession ID	OMIM Link
COL1A2	GDB:119062	COLLAGEN, TYPE I, ALPHA-2 POLYPEPTIDE; COL1A2 OSTEOGENESIS IMPERFECTA TYPE I OSTEOGENESIS IMPERFECTA TYPE IV; OI4
CRS	GDB:119073	CRANIOSYNOSTOSIS, TYPE 1; CRS1
CYMD	GDB:366594	MACULAR EDEMA, CYSTOID
DFNAs	GDB:636174	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 5; DFNAs
DLD	GDB:120608	LIPOAAMIDE DEHYDROGENASE DEFICIENCY, LACTIC ACIDOSIS DUE TO
DYT11	GDB:10013754	MYOCLONUS, HEREDITARY ESSENTIAL
EEC1	GDB:136338	ECTRODACTYLY, ECTODERMAL DYSPLASIA, AND CLEFT LIP/PALATE; EEC
ELN	GDB:119107	ELASTIN; ELN WILLIAMS-BEUREN SYNDROME; WBS
ETV1	GDB:335229	ETS VARIANT GENE 1; ETV1
FKBP6	GDB:9955215	WILLIAMS-BEUREN SYNDROME; WBS
GCK	GDB:127550	DIABETES MELLITUS, AUTOSOMAL DOMINANT, TYPE II GLUCOKINASE; GCK
GHRHR	GDB:138465	GROWTH HORMONE-RELEASING HORMONE RECEPTOR; GHRHR
GHS	GDB:9956363	MICROSOMIA WITH RADIAL DEFECTS
GLI3	GDB:119990	PALLISTER-HALL SYNDROME; PHS GLI-KRUPPEL FAMILY MEMBER 3; GLI3 POSTAXIAL POLYDACTYLY, TYPE A1 GREIG CEPHALOPOLYSYNDACTYLY SYNDROME; GCPS
GPDS1	GDB:9956410	GLAUCOMA, PIGMENT-DISPERSION TYPE
GUSB	GDB:120025	MUCOPOLYSACCHARIDOSIS TYPE VII

Gene	GDB Accession ID	OMIM Link
HADH	GDB:120033	HYDROXYACYL-CoA DEHYDROGENASE/3-KETOACYL-CoA THIOLASE/ENOYL-CoA HYDRATASE,
5 HLXB9	GDB:136411	HOME BOX GENE HB9; HLXB9 SACRAL AGENESIS, HEREDITARY, WITH PRESACRAL MASS, ANTERIOR MENINGOCELE,
HOXA13	GDB:120656	HOME BOX A13; HOXA13
10 HPFH2	GDB:128071	HEREDITARY PERSISTENCE OF FETAL HEMOGLOBIN, HETEROCELLULAR, INDIAN
HRX	GDB:9958999	HRX
IAB	GDB:11498909	ANEURYSM, INTRACRANIAL BERRY
15 IMMP2L	GDB:11499195	GILLES DE LA TOURETTE SYNDROME; GTS
KCNH2	GDB:138126	LONG QT SYNDROME, TYPE 2; LQT2
LAMB1	GDB:119357	LAMININ BETA 1; LAMB1
20 LEP	GDB:136420	LEPTIN; LEP
MET	GDB:120178	MET PROTO-ONCOGENE; MET
NCF1	GDB:120222	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-POSITIVE FORM
25 NM	GDB:119454	NEUTROPHIL CHEMOTACTIC RESPONSE; NCR
OGDH	GDB:118847	ALPHA-KETOGLUTARATE DEHYDROGENASE DEFICIENCY
OPN1SW	GDB:119032	TRITANOPIA
30 PEX1	GDB:9787110	ZELLWEGER SYNDROME; ZS PEROXIN-1; PEX1
PGAM2	GDB:120280	PHOSPHOGLYCERATE MUTASE, DEFICIENCY OF M SUBUNIT OF
PMS2	GDB:386406	POSTMEIOTIC SEGREGATION INCREASED (S. CEREVIAE)-2; PMS2
35 PON1	GDB:120308	PARAOXONASE 1; PON1

Gene	GDB Accession ID	OMIM Link
PPP1R3A	GDB:136797	PROTEIN PHOSPHATASE 1, REGULATORY (INHIBITOR) SUBUNIT 3; PPP1R3
PRSS1	GDB:119620	PANCREATITIS, HEREDITARY; PCTT PROTEASE, SERINE, 1; PRSS1
PTC	GDB:118744	PHENYLTHIOCARBAMIDE TASTING
PTPN12	GDB:136846	PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR TYPE, 12; PTPN12
RP10	GDB:138786	RETINITIS PIGMENTOSA-10; RP10
RP9	GDB:333931	RETINITIS PIGMENTOSA-9; RP9
SERPINE1	GDB:120297	PLASMINOGEN ACTIVATOR INHIBITOR, TYPE I; PAI1
SGCE	GDB:9958714	MYOCLONUS, HEREDITARY ESSENTIAL
SHFM1	GDB:128195	SPLIT-HAND/FOOT DEFORMITY, TYPE I; SHFD1
SHH	GDB:456309	HOLOPROSENCEPHALY, TYPE 3; HPE3 SONIC HEDGEHOG, DROSOPHILA, HOMOLOG OF; SHH
SLC26A3	GDB:138165	DOWN-REGULATED IN ADENOMA; DRA CHLORIDE DIARRHEA, FAMILIAL; CLD
SLC26A4	GDB:5584511	PENDRED SYNDROME; PDS DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 4; DFNB4
SLOS	GDB:385950	SMITH-LEMLI-OPITZ SYNDROME
SMAD1	GDB:3763345	SPINAL MUSCULAR ATROPHY, DISTAL, WITH UPPER LIMB PREDOMINANCE; SMAD1
TBXAS1	GDB:128744	THROMBOXANE A SYNTHASE 1; TBXAS1
TWIST	GDB:135694	ACROCEPHALOSYNDACTYLY TYPE III TWIST, DROSOPHILA, HOMOLOG OF; TWIST
ZWS1	GDB:120511	ZELLWEGER SYNDROME; ZS

Table 9: Genes, Locations and Genetic Disorders on Chromosome 8

Gene	GDB AccessionID	OMIM Link
5 ACHM3	GDB:9120558	PINGELAPESE BLINDNESS
ADRB3	GDB:203869	BETA-3-ADRENERGIC RECEPTOR; ADRB3
ANK1	GDB:118737	SPHEROCYTOSIS, HEREDITARY; HS
10 CA1	GDB:119047	CARBONIC ANHYDRASE I, ERYTHROCYTE, ELECTROPHORETIC VARIANTS OF; CA1
CA2	GDB:119739	OSTEOPETROSIS WITH RENAL TUBULAR ACIDOSIS
15 CCAL1	GDB:512892	CHONDROCALCINOSIS WITH EARLY-ONSET OSTEOARTHRITIS; CCAL2
CLN8	GDB:252118	EPILEPSY, PROGRESSIVE, WITH MENTAL RETARDATION; EPMR
15 CMT4A	GDB:138755	CHARCOT-MARIE-TOOTH NEUROPATHY 4A; CMT4A
CNGB3	GDB:9993286	PINGELAPESE BLINDNESS
COH1	GDB:252122	COHEN SYNDROME; COH1
20 CPP	GDB:119798	CERULOPLASMIN; CP
CRH	GDB:119804	CORTICOTROPIN-RELEASING HORMONE; CRH
25 CYP11B1	GDB:120603	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 11-@BETA-HYDROXYLASE DEFICIENCY
CYP11B2	GDB:120514	CYTOCHROME P450, SUBFAMILY XIB, POLYPEPTIDE 2; CYP11B2
DECR1	GDB:453934	2,4-@DIENOYL-CoA REDUCTASE; DECR
DPYS	GDB:5885803	DIHYDROPYRIMIDINASE; DPYS
30 DURS1	GDB:9958126	DUANE SYNDROME
EBS1	GDB:119856	EPIDERMOLYSIS BULLOSA SIMPLEX, OGNA TYPE
ECA1	GDB:10796318	JUVENILE ABSENCE
35 EGI	GDB:128830	EPILEPSY, GENERALIZED, IDIOPATHIC; EGI

	Gene	GDB AccessionID	OMIM Link
5	EXT1	GDB:135994	EXOSTOSES, MULTIPLE, TYPE I; EXT1 CHONDROSARCOMA
	EYA1	GDB:5215167	BRANCHIOOTORENAL DYSPLASIA EYES ABSENT 1; EYA1
	FGFR1	GDB:119913	ACROCEPHALOSYNDACTYLY TYPE V FIBROBLAST GROWTH FACTOR RECEPTOR-1; FGFR1
10	GNRH1	GDB:133746	GONADOTROPIN-RELEASING HORMONE 1; GNRH1 FAMILIAL HYPOGONADOTROPHIC
	GSR	GDB:119288	GLUTATHIONE REDUCTASE; GSR
15	GULOP	GDB:128078	SCURVY
	HR	GDB:595499	ALOPECIA UNIVERSALIS ATRICHLIA WITH PAPULAR LESIONS HAIRLESS, MOUSE, HOMOLOG OF
20	KCNQ3	GDB:9787230	CONVULSIONS, BENIGN FAMILIAL NEONATAL, TYPE 2; BFNC2 POTASSIUM CHANNEL, VOLTAGE-GATED, SUBFAMILY Q, MEMBER 3
	KFM	GDB:265291	KLIPPEL-FEIL SYNDROME; KFS; KFM
	KWE	GDB:9315120	KERATOLYTIC WINTER ERYTHEMA
	LGCR	GDB:120698	LANGER-GIEDION SYNDROME; LGS
	LPL	GDB:120700	HYPERLIPOPROTEINEMIA, TYPE I
25	MCPH1	GDB:9834525	MICROCEPHALY; MCT
	MOS	GDB:119396	TRANSFORMATION GENE: ONCOGENE MOS
	MYC	GDB:120208	TRANSFORMATION GENE: ONCOGENE MYC; MYC
	NAT1	GDB:125364	ARYLAMIDE ACETYLASE 1; AAC1
	NAT2	GDB:125365	ISONIAZID INACTIVATION
30	NBS1	GDB:9598211	NIJMEGEN BREAKAGE SYNDROME
	PLAT	GDB:119496	PLASMINOGEN ACTIVATOR, TISSUE; PLAT
	PLEC1	GDB:4119073	EPIDERMOLYSIS BULLOSA SIMPLEX AND LIMB-GIRDLE MUSCULAR DYSTROPHY PLECTIN 1; PLEC1
35			

Gene	GDB AccessionID	OMIM Link
PRKDC	GDB:234702	SEVERE COMBINED IMMUNODEFICIENCY DISEASE-1; SCID1 PROTEIN KINASE, DNA-ACTIVATED, CATALYTIC SUBUNIT; PRKDC
PXMP3	GDB:131487	PEROXIN-2; PEX2 ZELLWEGER SYNDROME; ZS
RP1	GDB:120352	RETINITIS PIGMENTOSA-1; RP1
SCZD6	GDB:9864736	DISORDER-2; SCZD2
SFTPC	GDB:120373	PULMONARY SURFACTANT APOPROTEIN PSP-C
SGM1	GDB:135350	KLIPPEL-FEIL SYNDROME; KFS; KFM
SPG5A	GDB:250332	SPASTIC PARAPLEGIA-5A, AUTOSOMAL RECESSIVE; SPG5A
STAR	GDB:635457	STEROIDOGENIC ACUTE REGULATORY PROTEIN; STAR
TG	GDB:120434	THYROGLOBULIN; TG
TRPS1	GDB:594960	TRICORRHINOPHALANGEAL SYNDROME, TYPE I; TRPS1
TTPA	GDB:512364	VITAMIN E, FAMILIAL ISOLATED DEFICIENCY OF; VED TOCOPHEROL (ALPHA) TRANSFER PROTEIN; TTPA
VMD1	GDB:119631	MACULAR DYSTROPHY, ATYPICAL VITELLIFORM; VMD1
WRN	GDB:128446	WERNER SYNDROME; WRN

Table 10: Genes, Locations and Genetic Disorders on Chromosome 9

Gene	GDB AccessionID	OMIM Link
ABCA1	GDB:305294	ANALPHALIPOPROTEINEMIA ATP-BINDING CASSETTE 1; ABC1
ABL1	GDB:119640	ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1; ABL1
ABO	GDB:118956	ABO BLOOD GROUP; ABO
ADAMTS13	GDB:9956467	THROMBOCYTOPENIC PURPURA
AK1	GDB:119664	ADENYLYLATE KINASE-1; AK1

Gene	GDB AccessionID	OMIM Link
ALAD	GDB:119665	DELTA-AMINOLEVULINATE DEHYDRATASE; ALAD
5 ALDH1A1	GDB:119667	ALDEHYDE DEHYDROGENASE-1; ALDH1
ALDOB	GDB:119669	FRUCTOSE INTOLERANCE, HEREDITARY
AMB P	GDB:120696	PROTEIN HC; HCP
10 AMCD1	GDB:437519	ARTHROGRYPOSIS MULTIPLEX CONGENITA, DISTAL, TYPE 1; AMCD1
ASS	GDB:119010	CITRULLINEMIA
BDMF	GDB:9954424	BONE DYSPLASIA WITH MEDULLARY FIBROSARCOMA
BSCL	GDB:9957720	SEIP SYNDROME
15 C5	GDB:119734	COMPLEMENT COMPONENT-5, DEFICIENCY OF
CDKN2A	GDB:335362	MELANOMA, CUTANEOUS MALIGNANT, 2; CMM2 CYCLIN-DEPENDENT KINASE INHIBITOR 2A; CDKN2A
20 CHAC	GDB:6268491	CHOREOACANTHOCYTOSIS; CHAC
CHH	GDB:138268	CARTILAGE-HAIR HYPOPLASIA; CHH
CMD1B	GDB:677147	CARDIOMYOPATHY, DILATED 1B; CMD1B
COL5A1	GDB:131457	COLLAGEN, TYPE V, ALPHA-1 POLYPEPTIDE; COL5A1
25 CRAT	GDB:359759	CARNITINE ACETYLTRANSFERASE; CRAT
DBH	GDB:119836	DOPAMINE BETA-HYDROXYLASE, PLASMA; DBH
DFNB11	GDB:1220180	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 7; DFNB7
30 DFNB7	GDB:636178	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 7; DFNB7
DNAI1	GDB:11500297	IMMOTILE CILIA SYNDROME-1; ICS1
DYS	GDB:137085	DYSAUTONOMIA, FAMILIAL; DYS
DYT1	GDB:119854	DYSTONIA 1, TORSION; DYT1
35 ENG	GDB:137193	ENDOGLIN; ENG

	Gene	GDB AccessionID	OMIM Link
5	EPB72	GDB:128993	ERYTHROCYTE SURFACE PROTEIN BAND 7.2; EPB72 STOMATOCYTOSIS I
	FANCC	GDB:132672	FANCONI ANEMIA, COMPLEMENTATION GROUP C; FACC
	FBP1	GDB:141539	FRUCTOSE-1,6-BISPHOPHATASE 1; FBP1
10	FCMD	GDB:250412	FUKUYAMA-TYPE CONGENITAL MUSCULAR DYSTROPHY; FCMD
	FRDA	GDB:119951	FRIEDREICH ATAXIA 1; FRDA1
	GALT	GDB:119971	GALACTOSEMIA
15	GLDC	GDB:128611	HYPERGLYCEMIA, ISOLATED NONKETOTIC, TYPE I; NKH1
	GNE	GDB:9954891	INCLUSION BODY MYOPATHY; IBM2
	GSM1	GDB:9784210	GENIOSPASM 1; GSM1
20	GSN	GDB:120019	AMYLOIDOSIS V GELSOLIN; GSN
	HSD17B3	GDB:347487	PSEUDOHERMAPHRODITISM, MALE, WITH GYNECOMASTIA
	HSN1	GDB:3853677	NEUROPATHY, HEREDITARY SENSORY, TYPE 1
25	IBM2	GDB:3801447	INCLUSION BODY MYOPATHY; IBM2
	LALL	GDB:9954426	LEUKEMIA, ACUTE, WITH LYMPHOMATOUS FEATURES; LALL
	LCCS	GDB:386141	LETHAL CONGENITAL CONTRACTURE SYNDROME; LCCS
30	LGMD2H	GDB:9862233	DYSTROPHY, HUTTERITE TYPE
	LMX1B	GDB:9834526	NAIL-PATELLA SYNDROME; NPS1
	MLLT3	GDB:138172	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 3; MLLT3
35	MROS	GDB:9954430	MELKERSSON SYNDROME
	MSSE	GDB:128019	EPITHELIOMA, SELF-HEALING SQUAMOUS
	NOTCH1	GDB:131400	NOTCH, DROSOPHILA, HOMOLOG OF, 1; NOTCH1
	ORM1	GDB:120250	OROSOMUCOID 1; ORM1

Gene	GDB Accession ID	OMIM Link
PAPPA	GDB:134729	PREGNANCY-ASSOCIATED PLASMA PROTEIN A; PAPPA
5	PIP5K1B	FRIEDREICH ATAXIA 1; FRDA1
	PTCH	BASAL CELL NEVUS SYNDROME; BCNS PATCHED, DROSOPHILA, HOMOLOG OF; PTCH
	PTGS1	PROSTAGLANDIN-ENDOPEROXIDASE SYNTHASE 1; PTGS1
	RLN1	RELAXIN; RLN1
10	RLN2	RELAXIN, OVARIAN, OF PREGNANCY
	RMRP	MITOCHONDRIAL RNA-PROCESSING ENDORIBONUCLEASE, RNA COMPONENT OF; RMRP; CARTILAGE-HAIR HYPOPLASIA; CHH
	ROR2	BRACHYDACTYLY, TYPE B; BDB ROBINOW SYNDROME, RECESSIVE FORM NEUROTROPHIC TYROSINE KINASE, RECEPTOR-RELATED 2; NTRKR2
15	RPD1	RETINITIS PIGMENTOSA-DEAFNESS SYNDROME 1, AUTOSOMAL DOMINANT
	SARDH	SARCOSINEMIA
	TDFA	FACTOR, AUTOSOMAL
20	TEK	VENOUS MALFORMATIONS, MULTIPLE CUTANEOUS AND MUCOSAL; VMCM TEK TYROSINE KINASE, ENDOTHELIAL; TEK
	TSC1	TUBEROUS SCLEROSIS-1; TSC1
	TYRP1	TYROSINASE-RELATED PROTEIN 1; TYRP1 ALBINISM III XANTHISM
25	XPA	XERODERMA PIGMENTOSUM I
30		

Table 11: Genes, Locations and Genetic Disorders on Chromosome 10

Gene	GDB Accession ID	OMIM Link
CACNB2	GDB:132014	CALCIUM CHANNEL, VOLTAGE-DEPENDENT, BETA-2 SUBUNIT; CACNB2

	Gene	GDB Accession ID	OMIM Link
5	COL17A1	GDB:131396	COLLAGEN, TYPE XVII, ALPHA-1 POLYPEPTIDE; COL17A1
	CUBN	GDB:636049	MEGALOBLASTIC ANEMIA 1; MGA1
	CYP17	GDB:119829	ADRENAL HYPERPLASIA, CONGENITAL, DUE TO 17-ALPHA-HYDROXYLASE DEFICIENCY
10	CYP2C19	GDB:119831	CYTOCHROME P450, SUBFAMILY IIC, POLYPEPTIDE 19; CYP2C19
	CYP2C9	GDB:131455	CYTOCHROME P450, SUBFAMILY IIC, POLYPEPTIDE 9; CYP2C9
15	EGR2	GDB:120611	EARLY GROWTH RESPONSE-2; EGR2
	EMX2	GDB:277886	EMPTY SPIRACLES, DROSOPHILA, 2, HOMOLOG OF; EMX2
	EPT	GDB:9786112	EPILEPSY, PARTIAL; EPT
20	ERCC6	GDB:119882	EXCISION-REPAIR CROSS-COMPLEMENTING RODENT REPAIR DEFICIENCY, COMPLEMENTATION
	FGFR2	GDB:127273	ACROCEPHALOSYNDACTYLY TYPE V FIBROBLAST GROWTH FACTOR RECEPTOR-2; FGFR2
25	HK1	GDB:120044	HEXOKINASE-1; HK1
	HOX11	GDB:119607	HOMEBOX-11; HOX11
	HPS	GDB:127359	HERMANSKY-PUDLAK SYNDROME; HPS
30	IL2RA	GDB:119345	INTERLEUKIN-2 RECEPTOR, ALPHA; IL2RA
	LGI1	GDB:9864936	EPILEPSY, PARTIAL; EPT
	LIPA	GDB:120153	WOLMAN DISEASE
35	MAT1A	GDB:129077	METHIONINE ADENOSYLTRANSFERASE DEFICIENCY
	MBL2	GDB:120167	MANNOSE-BINDING PROTEIN, SERUM; MBP1

	Gene	GDB Accession ID	OMIM Link
5	MKI67	GDB:120185	PROLIFERATION-RELATED Ki-67 ANTIGEN; MKI67
	MXI1	GDB:137182	MAX INTERACTING PROTEIN 1; MXI1
	OAT	GDB:120246	ORNITHINE AMINOTRANSFERASE DEFICIENCY
10	OATL3	GDB:215803	ORNITHINE AMINOTRANSFERASE DEFICIENCY
	PAX2	GDB:138771	PAIRED BOX HOMEOTIC GENE 2; PAX2
15	PCBD	GDB:138478	PTERIN-4-ALPHA-CARBINOLAMINE DEHYDRATASE; PCBD PRIMAPTERINURIA
	PEO1	GDB:632784	PROGRESSIVE EXTERNAL OPHTHALMOPLEGIA; PEO
	PHYH	GDB:9263423	REFSUM DISEASE PHYTANOYL-CoA HYDROXYLASE; PHYH
20	PNLIP	GDB:127916	LIPASE, CONGENITAL ABSENCE OF PANCREATIC
	PSAP	GDB:120366	PROSAPOSIN; PSAP
25	PTEN	GDB:6022948	MACROCEPHALY, MULTIPLE LIPOSAS AND HEMANGIOMATA MULTIPLE HAMARTOMA SYNDROME; MHAM POLYPOSIS, JUVENILE INTESTINAL PHOSPHATASE AND TENSIN HOMOLOG; PTEN
	RBP4	GDB:120342	RETINOL-BINDING PROTEIN, PLASMA; RBP4
30	RDPA	GDB:9954445	REFSUM DISEASE WITH INCREASED PIPECOLICACIDEMIA; RDPA
	RET	GDB:120346	RET PROTO-ONCOGENE; RET
	SDF1	GDB:433267	STROMAL CELL-DERIVED FACTOR 1; SDF1

Gene	GDB Accession ID	OMIM Link
SFTPA1	GDB:119593	PULMONARY SURFACTANT APOPROTEIN PSP-A; PSAP
SFTPD	GDB:132674	PULMONARY SURFACTANT APOPROTEIN PSP-D; PSP-D
SHFM3	GDB:386030	SPLIT-HAND/FOOT MALFORMATION, TYPE 3; SHFM3
SLAL	GDB:6549924	NEURAMINIDASE DEFICIENCY
THC2	GDB:10794765	THROMBOCYTOPENIA
TNFRSF6	GDB:132671	APOPTOSIS ANTIGEN 1; APT1
UFS	GDB:6380714	UROFACIAL SYNDROME; UFS
UROS	GDB:128112	PORPHYRIA, CONGENITAL ERYTHROPOIETIC; CEP

15 Table 12: Genes, Locations and Genetic Disorders on Chromosome 11

Gene	GDB Accession ID	OMIM Link
AA	GDB:568984	ATROPHIA AREATA; AA
ABCC8	GDB:591370	SULFONYLUREA RECEPTOR; SUR PERSISTENT HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY
ACAT1	GDB:126861	ALPHA-METHYLACETOACETICACIDURIA
ALX4	GDB:10450304	PARIETAL FORAMINA, SYMMETRIC; PFM
AMPD3	GDB:136013	ADENOSINE MONOPHOSPHATE DEAMINASE-3; AMPD3
ANC	GDB:9954484	CANAL CARCINOMA
APOA1	GDB:119684	AMYLOIDOSIS, FAMILIAL VISCERAL APOLIPOPROTEIN A-I OF HIGH DENSITY LIPOPROTEIN; APOA1
APOA4	GDB:119000	APOLIPOPROTEIN A-IV; APOA4
APOC3	GDB:119001	APOLIPOPROTEIN C-III; APOC3
ATM	GDB:593364	ATAXIA-TELANGIECTASIA; AT
BSCL2	GDB:9963996	SEIP SYNDROME
BWS	GDB:120567	BECKWITH-WIEDEMANN SYNDROME; BWS

	Gene	GDB Accession ID	OMIM Link
	CALCA	GDB:120571	CALCITONIN/CALCITONIN-RELATED POLYPEPTIDE, ALPHA; CALCA
5	CAT	GDB:119049	CATALASE; CAT
	CCND1	GDB:128222	LEUKEMIA, CHRONIC LYMPHATIC; CLL CYCLIN D1; CCND1
	CD3E	GDB:119764	CD3E ANTIGEN, EPSILON POLYPEPTIDE; CD3E
10	CD3G	GDB:119765	T3 T-CELL ANTIGEN, GAMMA CHAIN; T3G; CD3G
	CD59	GDB:119769	CD59 ANTIGEN P18-20; CD59 HUMAN LEUKOCYTE ANTIGEN MIC11; MIC11
	CDKN1C	GDB:593296	CYCLIN-DEPENDENT KINASE INHIBITOR 1C; CDKN1C
15	CLN2	GDB:125228	CEROID-LIPOFUSCINOSIS, NEURONAL 2, LATE INFANTILE TYPE; CLN2
	CNTF	GDB:125919	CILIARY NEUROTROPHIC FACTOR; CNTF
20	CPT1A	GDB:597642	HYPOGLYCEMIA, HYPOKETOTIC, WITH DEFICIENCY OF CARNITINE PALMITOYLTRANSFERASE CARNITINE PALMITOYLTRANSFERASE I, LIVER; CPT1A
25	CTSC	GDB:642234	KERATOSIS PALMOPLANTARIS WITH PERIODONTOPATHIA KERATOSIS PALMOPLANTARIS WITH PERIODONTOPATHIA AND ONYCHOGRYPOSIS CATHEPSIN C; CTSC
	DDB1	GDB:595014	DNA DAMAGE-BINDING PROTEIN; DDB1
	DDB2	GDB:595015	DNA DAMAGE-BINDING PROTEIN-2; DDB2
	DHCR7	GDB:9835302	SMITH-LEMLI-OPITZ SYNDROME
30	DLAT	GDB:118785	CIRRHOsis, PRIMARY; PBC
	DRD4	GDB:127782	DOPAMINE RECEPTOR D4; DRD4
	ECB2	GDB:9958955	POLYCYTHEMIA, BENIGN FAMILIAL
	ED4	GDB:9837373	DYSPLASIA, MARGARITA TYPE

Gene	GDB Accession ID	OMIM Link
EVR1	GDB:134029	EXUDATIVE VITREORETINOPATHY, FAMILIAL; EVR; EXT2; GDB:344921; EXOSTOSES, MULTIPLE, TYPE II; EXT2 CHONDROSARCOMA
F2	GDB:119894	COAGULATION FACTOR II; F2
FSHB	GDB:119955	FOLLICLE-STIMULATING HORMONE, BETA POLYPEPTIDE; FSHB
FTH1	GDB:120617	FERRITIN HEAVY CHAIN 1; FTH1
GIF	GDB:118800	PERNICIOUS ANEMIA, CONGENITAL, DUE TO DEFECT OF INTRINSIC FACTOR
GSD1B	GDB:9837619	GLYCOGEN STORAGE DISEASE Ib
GSD1C	GDB:9837637	STORAGE DISEASE Ic
HBB	GDB:119297	HEMOGLOBIN--BETA LOCUS; HBB
HBBP1	GDB:120035	HEMOGLOBIN--BETA LOCUS; HBB
HBD	GDB:119298	HEMOGLOBIN--DELTA LOCUS; HBD
HBE1	GDB:119299	HEMOGLOBIN--EPSILON LOCUS; HBE1
HBG1	GDB:119300	HEMOGLOBIN, GAMMA A; HBG1
HBG2	GDB:119301	HEMOGLOBIN, GAMMA G; HBG2
HMBS	GDB:120528	PORPHYRIA, ACUTE INTERMITTENT; AIP
HND	GDB:9954478	HARTNUP DISORDER
HOMG2	GDB:9956484	MAGNESIUM WASTING, RENAL
HRAS	GDB:120684	BLADDER CANCER V-HA-RAS HARVEY RAT SARCOMA VIRAL ONCOGENE HOMOLOG; HRAS
HVBS1	GDB:120069	CANCER, HEPATOCELLULAR
IDDM2	GDB:128530	DIABETES MELLITUS, INSULIN-DEPENDENT, 2 DIABETES MELLITUS, JUVENILE-ONSET INSULIN-DEPENDENT; IDDM
IGER	GDB:119696	IgE RESPONSIVENESS, ATOPIC; IGER
INS	GDB:119349	INSULIN; INS
JBS	GDB:120111	JACOBSEN SYNDROME; JBS

Gene	GDB Accession ID	OMIM Link
KCNJ11	GDB:7009893	POTASSIUM CHANNEL, INWARDLY-RECTIFYING, SUBFAMILY J, MEMBER 11; KCNJ11 PERSISTENT HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY
KCNJ1	GDB:204206	POTASSIUM CHANNEL, INWARDLY-RECTIFYING, SUBFAMILY J, MEMBER 1; KCNJ1
KCNQ1	GDB:741244	LONG QT SYNDROME, TYPE 1; LQT1
LDHA	GDB:120141	LACTATE DEHYDROGENASE-A; LDHA
LRP5	GDB:9836818	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME; OPPG HIGH BONE MASS
MEN1	GDB:120173	MULTIPLE ENDOCRINE NEOPLASIA, TYPE 1; MEN1
MLL	GDB:128819	MYELOID/LYMPHOID OR MIXED-LINEAGE LEUKEMIA; MLL
MTACR1	GDB:125743	MULTIPLE TUMOR ASSOCIATED CHROMOSOME REGION 1; MTACR1
MYBPC3	GDB:579615	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 4; CMH4 MYOSIN-BINDING PROTEIN C, CARDIAC; MYBPC3
MYO7A	GDB:132543	MYOSIN VIIA; MYO7A DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 2; DFNB2 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 11; DFNA11
NNO1	GDB:10450513	SIMPLE, AUTOSOMAL DOMINANT
OPPG	GDB:3789438	OSTEOPOROSIS-PSEUDOGLIOMA SYNDROME; OPPG
OPTB1	GDB:9954474	OSTEOPETROSIS, AUTOSOMAL RECESSIVE
PAX6	GDB:118997	PAIRED BOX HOMEOTIC GENE 6; PAX6
PC	GDB:119472	PYRUVATE CARBOXYLASE DEFICIENCY
PDX1	GDB:9836634	PYRUVATE DEHYDROGENASE COMPLEX, COMPONENT X

	Gene	GDB Accession ID	OMIM Link
5	PGL2	GDB:511177	PARAGANGLIOMAS, FAMILIAL NONCHROMAFFIN, 2; PGL2
	PGR	GDB:119493	PROGESTERONE RESISTANCE
	PORC	GDB:128610	PORPHYRIA, CHESTER TYPE; PORC
	PTH	GDB:119522	PARATHYROID HORMONE; PTH
10	PTS	GDB:118856	6-@PYRUVOYLtetrahydropterin SYNTHASE; PTS
	PVRL1	GDB:583951	ECTODERMAL DYSPLASIA, CLEFT LIP AND PALATE, HAND AND FOOT DEFORMITY, DYSPLASIA, MARGARITA TYPE POLIOVIRUS RECEPTOR RELATED; PVRR
15	PYGM	GDB:120329	GLYCOGEN STORAGE DISEASE V
	RAG1	GDB:120334	RECOMBINATION ACTIVATING GENE-1; RAG1
	RAG2	GDB:125186	RECOMBINATION ACTIVATING GENE-2; RAG2
20	ROM1	GDB:120350	ROD OUTER SEGMENT PROTEIN-1; ROM1
	SAA1	GDB:120364	SERUM AMYLOID A1; SAA1
25	SCA5	GDB:378219	SPINOCEREBELLAR ATAXIA 5; SCA5
	SCZD2	GDB:118874	DISORDER-2; SCZD2
30	SDHD	GDB:132456	PARAGANGLIOMAS, FAMILIAL NONCHROMAFFIN, 1; PGL1
	SERPING1	GDB:119041	ANGIONEUROTIC EDEMA, HEREDITARY; HANE
	SMPD1	GDB:128144	NIEMANN-PICK DISEASE
35	TCIRG1	GDB:9956269	OSTEOPETROSIS, AUTOSOMAL RECESSIVE
	TCL2	GDB:9954468	LEUKEMIA, ACUTE T-CELL; ATL
	TECTA	GDB:6837718	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 8; DFNA8 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 12; DFNA12

Gene	GDB Accession ID	OMIM Link
TH	GDB:119612	TYROSINE HYDROXYLASE; TH
TREH	GDB:9958953	TREHALASE
TSG101	GDB:1313414	TUMOR SUSCEPTIBILITY GENE 101; TSG101
TYR	GDB:120476	ALBINISM I
USH1C	GDB:132544	USHER SYNDROME, TYPE IC; USH1C
VMD2	GDB:133795	VITELLIFORM MACULAR DYSTROPHY; VMD2
VRNI	GDB:135662	VITREORETINOPATHY, NEOVASCULAR INFLAMMATORY; VRNI
WT1	GDB:120496	FRASIER SYNDROME WILMS TUMOR; WT1
WT2	GDB:118886	MULTIPLE TUMOR ASSOCIATED CHROMOSOME REGION 1; MTACR1
ZNF145	GDB:230064	PROMYELOCYTIC LEUKEMIA ZINC FINGER; PLZF

Table 13: Genes, Locations and Genetic Disorders on Chromosome 12

Gene	GDB Accession ID	OMIM Link
A2M	GDB:119639	ALPHA-2-MACROGLOBULIN; A2M
AAAS	GDB:9954498	GLUCOCORTICOID DEFICIENCY AND ACHALASIA
ACADS	GDB:118959	ACYL-CoA DEHYDROGENASE, SHORT-CHAIN; ACADS
ACLS	GDB:136346	ACROCALLOSAL SYNDROME; ACLS
ACVRL1	GDB:230240	OSLER-RENDU-WEBER SYNDROME 2; ORW2 ACTIVIN A RECEPTOR, TYPE II-LIKE KINASE 1; ACVRL1
ADHR	GDB:9954488	VITAMIN D-RESISTANT RICKETS, AUTOSOMAL DOMINANT
ALDH2	GDB:119668	ALDEHYDE DEHYDROGENASE-2; ALDH2
AMHR2	GDB:696210	ANTI-MULLERIAN HORMONE TYPE II RECEPTOR; AMHR2
AOM	GDB:118998	STICKLER SYNDROME, TYPE I; STL1

Gene	GDB Accession ID	OMIM Link
AQP2	GDB:141853	AQUAPORIN-2; AQP2 DIABETES INSIPIDUS, RENAL TYPE DIABETES INSIPIDUS, RENAL TYPE, AUTOSOMAL RECESSIVE
ATD	GDB:696353	ASPHYXIATING THORACIC DYSTROPHY; ATD
ATP2A2	GDB:119717	ATPase, Ca(2+)-TRANSPORTING, SLOW-TWITCH; ATP2A2 DARIER-WHITE DISEASE; DAR
BDC	GDB:5584359	BRACHYDACTYLY, TYPE C; BDC
C1R	GDB:119729	COMPLEMENT COMPONENT-C1r, DEFICIENCY OF
CD4	GDB:119767	T-CELL ANTIGEN T4/LEU3; CD4
CDK4	GDB:204022	CYCLIN-DEPENDENT KINASE 4; CDK4
CNA1	GDB:252119	CORNEA PLANA 1; CNA1
COL2A1	GDB:119063	STICKLER SYNDROME, TYPE I; STL1 COLLAGEN, TYPE II, ALPHA-1 CHAIN; COL2A1 ACHONDROGENESIS, TYPE II; ACG2
CYP27B1	GDB:9835730	PSEUDOVITAMIN D DEFICIENCY RICKETS; PDDR
DRPLA	GDB:270336	DENTATORUBRAL-PALLIDOLUYSIAN ATROPHY; DRPLA
ENUR2	GDB:666422	ENURESIS, NOCTURNAL, 2; ENUR2
FEOM1	GDB:345037	FIBROSIS OF EXTRAOCULAR MUSCLES, CONGENITAL; FEOM
FPP	GDB:9848880	PERIODIC FEVER, AUTOSOMAL DOMINANT
GNB3	GDB:120005	GUANINE NUCLEOTIDE-BINDING PROTEIN, BETA POLYPEPTIDE-3; GNB3
GNS	GDB:120006	MUCOPOLYSACCHARIDOSIS TYPE III
HAL	GDB:120746	HISTIDINEMIA

Gene	GDB Accession ID	OMIM Link
HBPI	GDB:701889	BRACHYDACTYLY WITH HYPERTENSION
5 HMGIC	GDB:362658	HIGH MOBILITY GROUP PROTEIN ISOFORM I-C; HMGIC
HMN2	GDB:9954508	MUSCULAR ATROPHY, ADULT SPINAL
HPD	GDB:135978	TYROSINEMIA, TYPE III
10 IGF1	GDB:120081	INSULINLIKE GROWTH FACTOR 1; IGF1
KCNA1	GDB:127903	POTASSIUM VOLTAGE-GATED CHANNEL, SHAKER-RELATED SUBFAMILY, MEMBER
15 KERA	GDB:252121	CORNEA PLANA 2; CNA2
KRAS2	GDB:120120	V-KI-RAS2 KIRSTEN RAT SARCOMA 2 VIRAL ONCOGENE HOMOLOG; KRAS2
KRT1	GDB:128198	KERATIN 1; KRT1
20 KRT2A	GDB:407640	ICHTHYOSIS, BULLOUS TYPE KERATIN 2A; KRT2A
KRT3	GDB:136276	KERATIN 3; KRT3
KRT4	GDB:120697	KERATIN 4; KRT4
25 KRT5	GDB:128110	EPIDERMOLYSIS BULLOSA HERPETIFORMIS, DOWLING-MEARA TYPE KERATIN 5; KRT5
KRT6A	GDB:128111	KERATIN 6A; KRT6A
KRT6B	GDB:128113	KERATIN 6B; KRT6B PACHYONYCHIA CONGENITA, JACKSON-LAWLER TYPE
30 KRTHB6	GDB:702078	MONILETHRIX KERATIN, HAIR BASIC (TYPE II) 6
LDHB	GDB:120147	LACTATE DEHYDROGENASE-B; LDHB
LYZ	GDB:120160	AMYLOIDOSIS, FAMILIAL VISCELAR LYSOZYME; LYZ
35 MGCT	GDB:9954504	TESTICULAR TUMORS

	Gene	GDB Accession ID	OMIM Link
	MPE	GDB:120191	MALIGNANT PROLIFERATION OF
5	MVK	GDB:134189	MEVALONICACIDURIA
	MYL2	GDB:128829	MYOSIN, LIGHT CHAIN, REGULATORY VENTRICULAR; MYL2
10	NS1	GDB:439388	NOONAN SYNDROME 1; NS1
	OAP	GDB:120245	OSTEOARTHROSIS, PRECOCIOUS; OAP
15	PAH	GDB:119470	PHENYLKETONURIA; PKU1
	PPKB	GDB:696352	PALMOPLANTAR KERATODERMA, BOTHNIAN TYPE; PPKB
20	PRB3	GDB:119513	PAROTID SALIVARY GLYCOPROTEIN; G1
	PXR1	GDB:433739	ZELLWEGER SYNDROME; ZS PEROXISOME RECEPTOR 1; PXR1
	RLS	GDB:11501392	ACROMELALGIA, HEREDITARY
25	RSN	GDB:139158	RESTIN; RSN
	SAS	GDB:128054	SARCOMA AMPLIFIED SEQUENCE; SAS
	SCA2	GDB:128034	SPINOCEREBELLAR ATAXIA 2; SCA2 ATAXIN-2; ATX2
	SCNN1A	GDB:366596	SODIUM CHANNEL, NONVOLTAGE-GATED, 1; SCNN1A
30	SMAL	GDB:9954506	SPINAL MUSCULAR ATROPHY, CONGENITAL NONPROGRESSIVE, OF LOWER LIMBS
	SPPM	GDB:9954502	SCAPULOPERONEAL MYOPATHY; SPM
	SPSMA	GDB:9954510	SCAPULOPERONEAL AMYOTROPHY, NEUROGENIC, NEW ENGLAND TYPE
	TBX3	GDB:681969	ULNAR-MAMMARY SYNDROME; UMS T-BOX 3; TBX3
35	TBX5	GDB:6175917	HOLT-ORAM SYNDROME; HOS T-BOX 5; TBX5

Gene	GDB Accession ID	OMIM Link
TCF1	GDB:125297	TRANSCRIPTION FACTOR 1, HEPATIC; TCF1 MATURITY-ONSET DIABETES OF THE YOUNG, TYPE III; MODY3
5 TPI1	GDB:119617	TRIOSEPHOSPHATE ISOMERASE 1; TPI1
TSC3	GDB:127930	SCLEROSIS-3; TSC3
10 ULR	GDB:594089	UTERINE
VDR	GDB:120487	VITAMIN D-RESISTANT RICKETS WITH END-ORGAN UNRESPONSIVENESS TO 1,25-DIHYDROXYCHOLECALCIFEROL VITAMIN D RECEPTOR; VDR
VWF	GDB:119125	VON WILLEBRAND DISEASE; VWD

15 Table 14: Genes, Locations and Genetic Disorders on Chromosome 13

Gene	GDB Accession ID	OMIM Link
ATP7B	GDB:120494	WILSON DISEASE; WND
20 BRCA2	GDB:387848	BREAST CANCER 2, EARLY-ONSET; BRCA2
BRCD1	GDB:9954522	BREAST CANCER, DUCTAL, 1; BRCD1
CLN5	GDB:230991	CEROID-LIPOFUSCOSIS, NEURONAL 5; CLN5
CPB2	GDB:129546	CARBOXYPEPTIDASE B2, PLASMA; CPB2
25 ED2	GDB:9834522	ECTODERMAL DYSPLASIA, HIDROTIC; HED
EDNRB	GDB:129075	ENDOTHELIN-B RECEPTOR; EDNRB HIRSCHSPRUNG DISEASE-2; HSCR2
ENUR1	GDB:594516	ENURESIS, NOCTURNAL, 1; ENUR1
30 ERCC5	GDB:120515	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 5; ERCC5
F10	GDB:119890	X, QUANTITATIVE VARIATION IN FACTOR X DEFICIENCY; F10
35 F7	GDB:119897	FACTOR VII DEFICIENCY

Gene	GDB Accession ID	OMIM Link
GJB2	GDB:125247	GAP JUNCTION PROTEIN, BETA-2, 26 KD; GJB2 DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 1; DFNB1 DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 3; DFNA3
GJB6	GDB:9958357	ECTODERMAL DYSPLASIA, HIDROTIC; HED DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 3; DFNA3
IPF1	GDB:448899	INSULIN PROMOTER FACTOR 1; IPF1
MBS1	GDB:128365	MOEBIUS SYNDROME; MBS
MCOR	GDB:9954520	CONGENITAL
PCCA	GDB:119473	GLYCINEMIA, KETOTIC, I
RB1	GDB:118734	BLADDER CANCER RETINOBLASTOMA; RB1
RHOK	GDB:371598	RHODOPSIN KINASE; RHOK
SCZD7	GDB:9864734	DISORDER-2; SCZD2
SGCG	GDB:3763329	MUSCULAR DYSTROPHY, LIMB GIRDLE, TYPE 2C; LGMD2C
SLC10A2	GDB:677534	SOLUTE CARRIER FAMILY 10, MEMBER 2; SLC10A2
SLC25A15	GDB:120042	HYPORORNITHINEMIA-HYPERAMMONEMIA-HOMOCITRULLINURIA SYNDROME
STARP1	GDB:635459	STEROIDOGENIC ACUTE REGULATORY PROTEIN; STAR
ZNF198	GDB:6382650	ZINC FINGER PROTEIN-198; ZNF198

Table 15: Genes, Locations and Genetic Disorders on Chromosome 14

Gene	GDB Accession ID	OMIM Link
ACHM1	GDB:132458	COLORBLINDNESS, TOTAL
ARVD1	GDB:371339	ARRHYTHMOGENIC RIGHT VENTRICULAR DYSPLASIA, FAMILIAL, 1; ARVD1
CTAA1	GDB:265299	CATARACT, ANTERIOR POLAR 1; CTAA1

Gene	GDB Accession ID	OMIM Link
DAD1	GDB:407505	DEFENDER AGAINST CELL DEATH; DAD1
DFNB5	GDB:636176	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 5; DFNB5
EML1	GDB:6328385	USHER SYNDROME, TYPE IA; USH1A
GALC	GDB:119970	KRABBE DISEASE
GCH1	GDB:118798	DYSTONIA, PROGRESSIVE, WITH DIURNAL VARIATION GTP CYCLOHYDROLASE I DEFICIENCY GTP CYCLOHYDROLASE I; GCH1
HE1	GDB:9957680	MALFORMATIONS, MULTIPLE, WITH LIMB ABNORMALITIES AND HYPOPITUITARISM
IBGC1	GDB:10450404	CEREBRAL CALCIFICATION, NONARTERIOSCLEROTIC
IGH@	GDB:118731	IgA CONSTANT HEAVY CHAIN 1; IGHA1 IMMUNOGLOBULIN: D (DIVERSITY) REGION OF HEAVY CHAIN IgA CONSTANT HEAVY CHAIN 2; IGHA2 IMMUNOGLOBULIN: J (JOINING) LOCI OF HEAVY CHAIN; IGHJ IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1 IMMUNOGLOBULIN: VARIABLE REGION OF HEAVY CHAINS--Hv1; IGHV IgG HEAVY CHAIN LOCUS; IGHG1 IMMUNOGLOBULIN Gm-2; IGHG2 IMMUNOGLOBULIN Gm-3; IGHG3 IMMUNOGLOBULIN Gm-4; IGHG4 IMMUNOGLOBULIN: HEAVY DELTA CHAIN; IGHD IMMUNOGLOBULIN: HEAVY EPSILON CHAIN; IGHE
IGHC group	GDB:9992632	IgA CONSTANT HEAVY CHAIN 1; IGHA1 IgA CONSTANT HEAVY CHAIN 2; IGHA2 IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1 IgG HEAVY CHAIN LOCUS; IGHG1 IMMUNOGLOBULIN Gm-2; IGHG2 IMMUNOGLOBULIN Gm-3; IGHG3 IMMUNOGLOBULIN Gm-4; IGHG4 IMMUNOGLOBULIN: HEAVY DELTA CHAIN; IGHD IMMUNOGLOBULIN: HEAVY EPSILON CHAIN; IGHE
IGHG1	GDB:120085	IgG HEAVY CHAIN LOCUS; IGHG1
IGHM	GDB:120086	IMMUNOGLOBULIN: HEAVY Mu CHAIN; Mu1; IGHM1

	Gene	GDB Accession ID	OMIM Link
5	IGHR	GDB:9954529	G1(A1) SYNDROME
	IV	GDB:139274	INVERSUS VISCRUM
	LTBP2	GDB:453890	LATENT TRANSFORMING GROWTH FACTOR-BETA BINDING PROTEIN 2; LTBP2
	MCOP	GDB:9954527	MICROPHTHALMOS
	MJD	GDB:118840	MACHADO-JOSEPH DISEASE; MJD
10	MNG1	GDB:6540062	GOITER, MULTINODULAR 1; MNG1
	MPD1	GDB:230271	MYOPATHY, LATE DISTAL HEREDITARY
	MPS3C	GDB:9954532	MUCOPOLYSACCHARIDOSIS TYPE III C
	MYH6	GDB:120214	MYOSIN, HEAVY POLYPEPTIDE 6; MYH6
	MYH7	GDB:120215	MYOSIN, CARDIAC, HEAVY CHAIN, BETA; MYH7
15	NP	GDB:120239	NUCLEOSIDE PHOSPHORYLASE; NP
	PABPN1	GDB:567135	OCULOPHARYNGEAL MUSCULAR DYSTROPHY; OPMD OCULOPHARYNGEAL MUSCULAR DYSTROPHY, AUTOSOMAL RECESSIVE POLYADENYLATE-BINDING PROTEIN-2; PABP2
	PSEN1	GDB:135682	ALZHEIMER DISEASE, FAMILIAL, TYPE 3; AD3
	PYGL	GDB:120328	GLYCOGEN STORAGE DISEASE VI
	RPGRIPI	GDB:11498766	AMAUROSIS CONGENITA OF LEBER I
20	SERPINA1	GDB:120289	PROTEASE INHIBITOR 1; PI
	SERPINA3	GDB:118955	ALPHA-1-ANTICHYMOTRYPsin; AACT
	SERPINA6	GDB:127865	CORTICOSTEROID-BINDING GLOBULIN; CBG
	SLC7A7	GDB:9863033	DIBASICAMINOACIDURIA II
	SPG3A	GDB:230126	SPASTIC PARAPLEGIA-3, AUTOSOMAL DOMINANT; SPG3A
25	SPTB	GDB:119602	ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE HEREDITARY HEMOLYTIC SPECTRIN, BETA, ERYTHROCYTIC; SPTB
	TCL1A	GDB:250785	T-CELL LYMPHOMA OR LEUKEMIA

Gene	GDB Accession ID	OMIM Link
TCRAV17S1	GDB:642130	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA
TCRAV5S1	GDB:451966	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA
TGM1	GDB:125299	TRANSGLUTAMINASE 1; TGM1 ICHTHYOSIS CONGENITA
TITF1	GDB:132588	THYROID TRANSCRIPTION FACTOR 1; TITF1
TMIP	GDB:9954523	AND ULNA, DUPLICATION OF, WITH ABSENCE OF TIBIA AND RADIUS
TRA@	GDB:120404	T-CELL ANTIGEN RECEPTOR, ALPHA SUBUNIT; TCRA
TSHR	GDB:125313	THYROTROPIN, UNRESPONSIVENESS TO
USH1A	GDB:118885	USHER SYNDROME, TYPE IA; USH1A
VP	GDB:120492	PORPHYRIA VARIEGATA

Table 16: Genes, Locations and Genetic Disorders on Chromosome 15

Gene	GDB Accession ID	OMIM Link
ACCPN	GDB:5457725	CORPUS CALLOSUM, AGENESIS OF, WITH NEURONOPATHY
AHO2	GDB:9954535	HEREDITARY OSTEODYSTROPHY-2; AHO2
ANCR	GDB:119678	ANGELMAN SYNDROME
B2M	GDB:119028	BETA-2-MICROGLOBULIN; B2M
BBS4	GDB:511199	BARDET-BIEDL SYNDROME, TYPE 4; BBS4
BLM	GDB:135698	BLOOM SYNDROME; BLM
CAPN3	GDB:119751	CALPAIN, LARGE POLYPEPTIDE L3; CAPN3 MUSCULAR DYSTROPHY, LIMB-GIRDLE, TYPE 2; LGMD2
CDAN1	GDB:9823267	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE I
CDAN3	GDB:386192	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE III; CDAN3

35

Gene	GDB Accession ID	OMIM Link
CLN6	GDB:4073043	CERODD-LIPOFUSCINOSIS, NEURONAL 6, LATE INFANTILE, VARIANT; CLN6
5 CMH3	GDB:138299	CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 3; CMH3
CYP19	GDB:119830	CYTOCHROME P450, SUBFAMILY XIX; CYP19
10 CYP1A1	GDB:120604	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 1; CYP1A1
CYP1A2	GDB:118780	CYTOCHROME P450, SUBFAMILY I, POLYPEPTIDE 2; CYP1A2
DYX1	GDB:1391796	DYSLEXIA, SPECIFIC, 1; DYX1
15 EPB42	GDB:127385	HEREDITARY HEMOLYTIC PROTEIN 4.2, ERYTHROCYTIC; EPB42
ETFA	GDB:119121	GLUTARICACIDURIA IIA; GA IIA
EYCL3	GDB:4590306	EYE COLOR-3; EYCL3
FAH	GDB:119901	TYROSINEMIA, TYPE I
20 FBN1	GDB:127115	FIBRILLIN-1; FBN1 MARFAN SYNDROME; MFS
FES	GDB:119906	V-FES FELINE SARCOMA VIRAL/V-FPS FUJINAMI AVIAN SARCOMA VIRAL ONCOGENE
25 HCVS	GDB:119306	CORONAVIRUS 229E SUSCEPTIBILITY; CVS
HEXA	GDB:120040	TAY-SACHS DISEASE; TSD
IVD	GDB:119354	ISOVALERICACIDEMIA; IVA
LCS1	GDB:11500552	CHOLESTASIS-LYMPHEDEMA SYNDROME
30 LIPC	GDB:119366	LIPASE, HEPATIC; LIPC
MYO5A	GDB:218824	MYOSIN VA; MYO5A
OCA2	GDB:136820	ALBINISM II
OTSC1	GDB:9860473	OTOSCLEROSIS
35 PWCR	GDB:120325	PRADER-WILLI SYNDROME

Gene	GDB Accession ID	OMIM Link
5	RLBP1	RETINALDEHYDE-BINDING PROTEIN 1;; RLBP1
	SLC12A1	SOLUTE CARRIER FAMILY 12, MEMBER 1; SLC12A1
	SPG6	SPASTIC PARAPLEGIA 6, AUTOSOMAL DOMINANT; SPG6
	TPM1	TROPOMYOSIN 1; TPM1
	UBE3A	ANGELMAN SYNDROME UBIQUITIN-PROTEIN LIGASE E3A; UBE3A
10	WMS	WEILL-MARCHESANI SYNDROME

Table 17: Genes, Locations and Genetic Disorders on Chromosome 16

Gene	GDB Accession ID	OMIM Link
15	ABCC6	PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL DOMINANT; PXE PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL RECESSIVE; PXE
	ALDOA	ALDOLASE A, FRUCTOSE-BISPHOSPHATE; ALDOA
	APRT	ADENINE PHOSPHORIBOSYLTRANSFERASE; APRT
	ATP2A1	ATPase, Ca(2+)-TRANSPORTING, FAST-TWITCH 1; ATP2A1 BRODY MYOPATHY
	BBS2	BARDET-BIEDL SYNDROME, TYPE 2; BBS2
20	CARD15	SYNOVITIS, GRANULOMATOUS, WITH UVEITIS AND CRANIAL NEUROPATHIES REGIONAL ENTERITIS
25	CATM	MICROPHTHALMIA-CATARACT
30	CDH1	CADHERIN 1; CDH1
35	CETP	CHOLESTERYL ESTER TRANSFER PROTEIN, PLASMA; CETP
	CHST6	CORNEAL DYSTROPHY, MACULAR TYPE
	CLN3	CEROID-LIPOFUSCINOSIS, NEURONAL 3, JUVENILE; CLN3

	Gene	GDB Accession ID	OMIM Link
5	CREBBP	GDB:437159	RUBINSTEIN SYNDROME CREB-BINDING PROTEIN; CREBBP
	CTH	GDB:119086	CYSTATHIONINURIA
	CTM	GDB:119819	CATARACT, ZONULAR
	CYBA	GDB:125238	GRANULOMATOUS DISEASE, CHRONIC, AUTOSOMAL CYTOCHROME-b-NEGATIVE FORM
10	CYLD	GDB:701216	EPITHELIOMA, HEREDITARY MULTIPLE BENIGN CYSTIC
	DHS	GDB:9958268	XEROCYTOSIS, HEREDITARY
	DNASE1	GDB:132846	DEOXYRIBONUCLEASE I; DNASE1
	DPEP1	GDB:128059	RENAL DIPEPTIDASE
15	ERCC4	GDB:119113	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 4; ERCC4 XERODERMA PIGMENTOSUM, COMPLEMENTATION GROUP F; XPF
	FANCA	GDB:701221	FANCONI ANEMIA, COMPLEMENTATION GROUP A; FACA
	GALNS	GDB:129085	MUCOPOLYSACCHARIDOSIS TYPE IVA
	GAN	GDB:9864885	NEUROPATHY, GIANT AXONAL; GAN
20	HAGH	GDB:119292	HYDROXYACYL GLUTATHIONE HYDROLASE; HAGH
	HBA1	GDB:119293	HEMOGLOBIN-ALPHA LOCUS-1; HBA1
	HBA2	GDB:119294	HEMOGLOBIN-ALPHA LOCUS-2; HBA2
	HBHR	GDB:9954541	HEMOGLOBIN H-RELATED MENTAL RETARDATION
25	HBQ1	GDB:120036	HEMOGLOBIN-THETA-1 LOCUS; HBQ1
	HBZ	GDB:119302	HEMOGLOBIN-ZETA LOCUS; HBZ
	HBZP	GDB:120037	HEMOGLOBIN-ZETA LOCUS; HBZ
	HP	GDB:119314	HAPTOGLOBIN; HP
30	HSD11B2	GDB:409951	CORTISOL 11-BETA-KETOREDUCTASE DEFICIENCY
	IL4R	GDB:118823	INTERLEUKIN-4 RECEPTOR; IL4R

Gene	GDB Accession ID	OMIM Link
LIPB	GDB:119365	LIPASE B, LYSOSOMAL ACID; LIPB
5	MC1R	MELANOCORTIN-1 RECEPTOR; MC1R
	MEFV	MEDITERRANEAN FEVER, FAMILIAL; MEFV
	MHC2TA	MHC CLASS II TRANSACTIVATOR; MHC2TA
	MLYCD	MALONYL CoA DECARBOXYLASE DEFICIENCY
10	PHKB	PHOSPHORYLASE KINASE, BETA SUBUNIT; PHKB
	PHKG2	PHOSPHORYLASE KINASE, TESTIS/LIVER, GAMMA 2; PHKG2
	PKD1	POLYCYSTIC KIDNEYS POLYCYSTIC KIDNEY DISEASE 1; PKD1
15	PKDTS	POLYCYSTIC KIDNEY DISEASE, INFANTILE SEVERE, WITH TUBEROUS SCLEROSIS;
	PMM2	CARBOHYDRATE-DEFICIENT GLYCOPROTEIN SYNDROME, TYPE I; CDG1 PHOSPHOMANNOMUTASE 2; PMM2
	PXE	PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL DOMINANT; PXE PSEUDOXANTHOMA ELASTICUM, AUTOSOMAL RECESSIVE; PXE
20	SALL1	TOWNES-BROCKS SYNDROME; TBS SAL-LIKE 1; SALL1
	SCA4	SPINOCEREBELLAR ATAXIA 4; SCA4
25	SCNN1B	SODIUM CHANNEL, NONVOLTAGE-GATED 1 BETA; SCNN1B
	SCNN1G	SODIUM CHANNEL, NONVOLTAGE-GATED 1 GAMMA; SCNN1G
	TAT	TYROSINE TRANSAMINASE DEFICIENCY
30	TSC2	TUBEROUS SCLEROSIS-2; TSC2
	VDI	DEFECTIVE INTERFERING PARTICLE INDUCTION, CONTROL OF
	WT3	WILMS TUMOR, TYPE III; WT3

Table 18: Genes, Locations and Genetic Disorders on Chromosome 17

	Gene	GDB Accession ID	OMIM Link
5	ABR	GDB:119642	ACTIVE BCR-RELATED GENE; ABR
	ACACA	GDB:120534	ACETYL-CoA CARBOXYLASE DEFICIENCY
	ACADVL	GDB:1248185	ACYL-CoA DEHYDROGENASE, VERY-LONG-CHAIN, DEFICIENCY OF
10	ACE	GDB:119840	DIPEPTIDYL CARBOXYPEPTIDASE-1; DCP1
	ALDH3A2	GDB:1316855	SJOGREN-LARSSON SYNDROME; SLS
	APOH	GDB:118887	APOLIPOPROTEIN H; APOH
15	ASPA	GDB:231014	SPONGY DEGENERATION OF CENTRAL NERVOUS SYSTEM
	AXIN2	GDB:9864782	CANCER OF COLON
	BCL5	GDB:125178	LEUKEMIA/LYMPHOMA, CHRONIC B-CELL, 5; BCL5
20	BHD	GDB:11498904	WITH TRICHODISCOMAS AND ACROCHORDONS
	BLMH	GDB:3801467	BLEOMYCIN HYDROLASE
	BRCA1	GDB:126611	BREAST CANCER, TYPE 1; BRCA1
25	CACD	GDB:5885801	CHOROIDAL DYSTROPHY, CENTRAL AREOLAR; CACD
	CCA1	GDB:118763	CATARACT, CONGENITAL, CERULEAN TYPE 1; CCA1
	CCZS	GDB:681973	CATARACT, CONGENITAL ZONULAR, WITH SUTURAL OPACITIES; CCZS
30	CHRN1	GDB:120587	CHOLINERGIC RECEPTOR, NICOTINIC, BETA POLYPEPTIDE 1; CHRN1
	CHRNE	GDB:132246	CHOLINERGIC RECEPTOR, NICOTINIC, EPSILON POLYPEPTIDE; CHRNE
	CMT1A	GDB:119785	CHARCOT-MARIE-TOOTH DISEASE, TYPE 1A; CMT1A NEUROPATHY, HEREDITARY, WITH LIABILITY TO PRESSURE PALSIES; HNPP

Gene	GDB Accession ID	OMIM Link
COL1A1	GDB:119061	COLLAGEN, TYPE I, ALPHA-1 CHAIN; COL1A1 OSTEOPENESIS IMPERFECTA TYPE I OSTEOPENESIS IMPERFECTA TYPE IV; OI4
CORD5	GDB:568473	CONE-ROD DYSTROPHY-5; CORD5
CTNS	GDB:700761	CYSTINOSIS, EARLY-ONSET OR INFANTILE NEPHROPATHIC TYPE
EPX	GDB:377700	EOSINOPHIL PEROXIDASE; EPX
ERBB2	GDB:120613	V-ERB-B2 AVIAN ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE HOMOLOG 2; ERBB2
G6PC	GDB:231927	GLYCOGEN STORAGE DISEASE I; GSD-I
GAA	GDB:119965	GLYCOGEN STORAGE DISEASE II
GALK1	GDB:119246	GALACTOKINASE DEFICIENCY
GCGR	GDB:304516	GLUCAGON RECEPTOR; GCGR
GFAP	GDB:118799	GLIAL FIBRILLARY ACIDIC PROTEIN; GFAP ALEXANDER DISEASE
GH1	GDB:119982	GROWTH HORMONE 1; GH1
GH2	GDB:119983	GROWTH HORMONE 2; GH2
GP1BA	GDB:118806	GIANT PLATELET SYNDROME
GPSC	GDB:9954564	FAMILIAL PROGRESSIVE SUBCORTICAL
GUCY2D	GDB:136012	AMAUROSIS CONGENITA OF LEBER I GUANYLATE CYCLASE 2D, MEMBRANE; GUC2D CONE-ROD DYSTROPHY-6; CORD6
ITGA2B	GDB:120012	THROMBASTHENIA OF GLANZMANN AND NAEGELI
ITGB3	GDB:120013	INTEGRIN, BETA-3; ITGB3
ITGB4	GDB:128028	INTEGRIN, BETA-4; ITGB4
KRT10	GDB:118828	KERATIN 10; KRT10
KRT12	GDB:5583953	CORNEAL DYSTROPHY, JUVENILE EPITHELIAL, OF MEESMANN KERATIN 12; KRT12

Gene	GDB Accession ID	OMIM Link
KRT13	GDB:120740	KERATIN 13; KRT13
KRT14	GDB:132145	KERATIN 14; KRT14 GLUTATHIONE SYNTHETASE; GSS
KRT14L1	GDB:120121	KERATIN 14; KRT14
KRT14L2	GDB:120122	KERATIN 14; KRT14
KRT14L3	GDB:120123	KERATIN 14; KRT14
KRT16	GDB:136207	KERATIN 16; KRT16
KRT16L1	GDB:120125	KERATIN 16; KRT16
KRT16L2	GDB:120126	KERATIN 16; KRT16
KRT17	GDB:136211	KERATIN 17; KRT17 PACHYONYCHIA CONGENITA, JACKSON-LAWLER TYPE
KRT9	GDB:303970	HYPERKERATOSIS, LOCALIZED EPIDERMOLYTIC
MAPT	GDB:119434	MICROTUBULE-ASSOCIATED PROTEIN TAU; MAPT PALLIDOPONTONIGRAL DEGENERATION; PPND DISINHIBITION-DEMENTIA-PARKINSONI SM-AMYOTROPHY COMPLEX; DDPAC
MDB	GDB:9958959	MEDULLOBLASTOMA; MDB
MDCR	GDB:120525	MILLER-DIEKER LISSENCEPHALY SYNDROME; MDLS PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
MGI	GDB:9954550	MYASTHENIA GRAVIS, FAMILIAL INFANTILE; FIMG
MHS2	GDB:132580	MALIGNANT HYPERTHERMIA SUSCEPTIBILITY-2; MHS2
MKS1	GDB:681967	MECKEL SYNDROME; MKS
MPO	GDB:120192	MYELOPEROXIDASE DEFICIENCY
MUL	GDB:636050	MULIBREY NANISM; MUL
MYO15A	GDB:9838006	DEAFNESS, NEUROSENSORY, AUTOSOMAL RECESSIVE, 3; DFNB3
NAGLU	GDB:636533	MUCOPOLYSACCHARIDOSIS TYPE IIIB

	Gene	GDB Accession ID	OMIM Link
5	NAPB	GDB:9954572	NEURITIS WITH BRACHIAL PREDILECTION; NAPB
	NF1	GDB:120231	NEUROFIBROMATOSIS, TYPE I; NF1
	NME1	GDB:127965	NON-METASTATIC CELLS 1, PROTEIN EXPRESSED IN; NME1
10	P4HB	GDB:120708	PROLYL-4-HYDROXYLASE, BETA POLYPEPTIDE; PHDB; PROHB
	PAFAH1B1	GDB:677430	MILLER-DIEKER LISSENCEPHALY SYNDROME; MDLS PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE, GAMMA SUBUNIT
15	PECAM1	GDB:696372	PLATELET-ENDOTHELIAL CELL ADHESION MOLECULE; PECAM1
	PEX12	GDB:6155804	ZELLWEGER SYNDROME; ZS PEROXIN-12; PEX12
20	PHB	GDB:126600	PROHIBITIN; PHB
	PMP22	GDB:134190	CHARCOT-MARIE-TOOTH DISEASE, TYPE 1A; CMT1A HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS PERIPHERAL MYELIN PROTEIN 22; PMP22
25	PRKAR1A	GDB:120313	MYXOMA, SPOTTY PIGMENTATION, AND ENDOCRINE OVERACTIVITY PROTEIN KINASE, cAMP-DEPENDENT, REGULATORY, TYPE I, ALPHA; PRKAR1A
	PRKCA	GDB:128015	PROTEIN KINASE C, ALPHA; PRKCA
30	PRKWNK4	GDB:9954566	PSEUDOHYPOALDOSTERONISM TYPE II, LOCUS B; PHA2B
	PRP8	GDB:9957697	RETINITIS PIGMENTOSA-13; RP13
35	PRPF8	GDB:392647	RETINITIS PIGMENTOSA-13; RP13
	PTLAH	GDB:9957342	APLASIA OR HYPOPLASIA
	RARA	GDB:120337	RETINOIC ACID RECEPTOR, ALPHA; RARA
	RCV1	GDB:135477	RECOVERIN; RCV1

Gene	GDB Accession ID	OMIM Link
RMSA1	GDB:304519	REGULATOR OF MITOTIC SPINDLE ASSEMBLY 1; RMSA1
5	RP17	RETINITIS PIGMENTOSA-17; RP17
	RSS	RUSSELL-SILVER SYNDROME; RSS
	SCN4A	PERIODIC PARALYSIS II
	SERPINF2	PLASMIN INHIBITOR DEFICIENCY
10	SGCA	ADHALIN; ADL
	SGSH	MUCOPOLYSACCHARIDOSIS TYPE IIIA
	SHBG	SEX HORMONE BINDING GLOBULIN; SHBG
15	SLC2A4	SOLUTE CARRIER FAMILY 2, MEMBER 4; SLC2A4
	SLC4A1	SOLUTE CARRIER FAMILY 4, ANION EXCHANGER, MEMBER 1; SLC4A1 BLOOD GROUP--DIEGO SYSTEM; DI BLOOD GROUP--WRIGHT ANTIGEN; Wr ELLIPTOCYTOSIS, RHESUS-UNLINKED TYPE HEREDITARY HEMOLYTIC
20	SLC6A4	SOLUTE CARRIER FAMILY 6, MEMBER 4; SLC6A4
	SMCR	SMITH-MAGENIS SYNDROME; SMS
	SOST	SCLEROSTESIS
25	SOX9	DYSPLASIA
	SSTR2	SOMATOSTATIN RECEPTOR-2; SSTR2
30	SYM1	SYMPHALANGISM, PROXIMAL; SYM1
	SYNS1	SYNOSTOSES, MULTIPLE, WITH BRACHYDACTYLY
	TCF2	TRANSCRIPTION FACTOR-2, HEPATIC; TCF2
35	THRA	THYROID HORMONE RECEPTOR, ALPHA 1; THRA
	TIMP2	TISSUE INHIBITOR OF METALLOPROTEINASE-2; TIMP2

Gene	GDB Accession ID	OMIM Link
TOC	GDB:451978	TYLOSIS WITH ESOPHAGEAL CANCER; TOC
5 TOP2A	GDB:118884	TOPOISOMERASE (DNA) II, ALPHA; TOP2A
TP53	GDB:120445	CANCER, HEPATOCELLULAR LI-FRAUMENI SYNDROME; LFS TUMOR PROTEIN p53; TP53 CARCINOMA
10 VBCH	GDB:9954554	HYPEROSTOSIS CORTICALIS GENERALISATA

Table 19: Genes, Locations and Genetic Disorders on Chromosome 18

Gene	GDB Accession ID	OMIM Link
15 ATP8B1	GDB:453352	CHOLESTASIS, PROGRESSIVE FAMILIAL INTRAHEPATIC 1; PFIC1 INTRAHEPATIC CHOLESTASIS FAMILIAL INTRAHEPATIC CHOLESTASIS-1; FIC1
BCL2	GDB:119031	B-CELL CLL/LYMPHOMA 2; BCL2
CNSN	GDB:9954580	CARNOSINEMIA
20 CORD1	GDB:118773	CONE-ROD DYSTROPHY-1; CORD1
CYB5	GDB:125236	METHEMOGLOBINEMIA DUE TO DEFICIENCY OF CYTOCHROME b5
DCC	GDB:119838	DELETED IN COLORECTAL CARCINOMA; DCC
25 F5F8D	GDB:6919858	FACTOR V AND FACTOR VIII, COMBINED DEFICIENCY OF; F5F8D
FECH	GDB:127282	PROTOPORPHYRIA, ERYTHROPOIETIC
FEO	GDB:4378120	POLYOSTOTIC OSTEOLYTIC DYSPLASIA, HEREDITARY EXPANSILE; HEPOD
30 LAMA3	GDB:251818	LAMININ, ALPHA 3; LAMA3
LCFS2	GDB:9954578	CANCER
MADH4	GDB:4642788	POLYPOSIS, JUVENILE INTESTINAL MOTHERS AGAINST DECAPENTAPLEGIC, DROSOPHILA, HOMOLOG OF, 4; MADH4
35 MAFD1	GDB:120163	MANIC-DEPRESSIVE PSYCHOSIS, AUTOSOMAL
MC2R	GDB:135163	ADRENAL UNRESPONSIVENESS TO ACTH

Gene	GDB Accession ID	OMIM Link
MCL	GDB:9954574	LEIOMYOMATA, HEREDITARY MULTIPLE, OF SKIN
5	MYP2	MYOPIA
	NPC1	NIEMANN-PICK DISEASE, TYPE C1; NPC1
	SPPK	PALMOPLANTARIS STRIATA
	TGFBRE	TRANSFORMING GROWTH FACTOR, BETA 1 RESPONSE ELEMENT
	TGIF	HOLOPROSENCEPHALY, TYPE 4; HPE4
	TTR	TRANSTHYRETIN; TTR

Table 20: Genes, Locations and Genetic Disorders on Chromosome 19

Gene	GDB Accession ID	OMIM Link
15	AD2	ALZHEIMER DISEASE-2; AD2
	AMH	PERSISTENT MULLERIAN DUCT SYNDROME, TYPES I AND II; PMDS ANTI-MULLERIAN HORMONE; AMH
	APOC2	APOLIPOPROTEIN C-II DEFICIENCY, TYPE I HYPERLIPOPROTEINEMIA DUE TO
	APOE	APOLIPOPROTEIN E; APOE
	ATHS	LIPOPROTEIN PHENOTYPE; ALP
	BAX	BCL2-ASSOCIATED X PROTEIN; BAX
	BCKDHA	MAPLE SYRUP URINE DISEASE
	BCL3	B-CELL LEUKEMIA/LYMPHOMA-3; BCL3
	BFIC	BENIGN FAMILIAL INFANTILE CONVULSIONS
25	C3	COMPLEMENT COMPONENT-3; C3
	CACNA1A	ATAXIA, PERIODIC VESTIBULOCEREBELLAR HEMIPLEGIC MIGRAINE, FAMILIAL; MHP SPINOCEBELLAR ATAXIA 6; SCA6 CALCIUM CHANNEL, VOLTAGE-DEPENDENT, P/Q TYPE, ALPHA 1A SUBUNIT; CACNA1A
30		
35		

Gene	GDB Accession ID	OMIM Link
CCO	GDB:119755	CENTRAL CORE DISEASE OF MUSCLE
CEACAM5	GDB:119054	CARCINOEMBRYONIC ANTIGEN; CEA
COMP	GDB:344263	EPIPHYSEAL DYSPLASIA, MULTIPLE; MED PSEUDOACHONDROPLASTIC DYSPLASIA; CARTILAGE OLIGOMERIC MATRIX PROTEIN; COMP
CRX	GDB:333932	CONE-ROD DYSTROPHY-2; CORD2; AMAUROSIS CONGENITA OF LEBER I; CONE-ROD HOMEOPBOX-CONTAINING GENE
DBA	GDB:9600353	ANEMIA, CONGENITAL HYPOPLASTIC, OF BLACKFAN AND DIAMOND
DDU	GDB:10796026	URTICARIA; DDU
DFNA4	GDB:606540	DEAFNESS, AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL, 4; DFNA4
DLL3	GDB:9959026	VERTEBRAL ANOMALIES
DMPK	GDB:119097	DYSTROPHIA MYOTONICA; DM
DMWD	GDB:7178354	DYSTROPHIA MYOTONICA; DM
DPD1	GDB:10796170	ENGELMANN DISEASE
E11S	GDB:119101	ECHO 11 SENSITIVITY; E11S
ELA2	GDB:118792	ELASTASE-2; ELA2 NEUTROOPENIA, CYCLIC
EPOR	GDB:125242	ERYTHROPOIETIN RECEPTOR; EPOR
ERCC2	GDB:119112	EXCISION-REPAIR, COMPLEMENTING DEFECTIVE, IN CHINESE HAMSTER, 2; ERCC2 XERODERMA PIGMENTOSUM IV; XP4
ETFB	GDB:119887	ELECTRON TRANSFER FLAVOPROTEIN, BETA POLYPEPTIDE; ETFB
EXT3	GDB:383780	EXOSTOSES, MULTIPLE, TYPE III; EXT3
EYCL1	GDB:119269	EYE COLOR-1; EYCL1
FTL	GDB:119234	FERRITIN LIGHT CHAIN; FTL
FUT1	GDB:120618	FUCOSYLTRANSFERASE-1; FUT1

	Gene	GDB Accession ID	OMIM Link
	FUT2	GDB:120619	FUCOSYLTRANSFERASE-2; FUT2
5	FUT6	GDB:135180	FUCOSYLTRANSFERASE-6; FUT6
	GAMT	GDB:1313736	GUANIDINOACETATE METHYLTRANSFERASE; GAMT
	GCDH	GDB:136004	GLUTARICACIDEMIA I
	GPI	GDB:120015	GLUCOSEPHOSPHATE ISOMERASE; GPI
10	GUSM	GDB:119291	GLUCURONIDASE, MOUSE, MODIFIER OF; GUSM
	HB1	GDB:9954586	BUNDLE BRANCH BLOCK
	HCL1	GDB:119304	HAIR COLOR-1; HCL1
15	HHC2	GDB:249836	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL, TYPE II; HHC2
	HHC3	GDB:9955121	HYPOCALCIURIC HYPERCALCEMIA, FAMILIAL, TYPE III; HHC3
	ICAM3	GDB:136236	INTERCELLULAR ADHESION MOLECULE-3; ICAM3
20	INSR	GDB:119352	INSULIN RECEPTOR; INSR
	JAK3	GDB:376460	JANUS KINASE 3 JAK3
	KLK3	GDB:119695	ANTIGEN, PROSTATE-SPECIFIC; APS
	LDLR	GDB:119362	HYPERCHOLESTEROLEMIA, FAMILIAL; FHC
25	LHB	GDB:119364	LUTEINIZING HORMONE, BETA POLYPEPTIDE; LHB
	LIG1	GDB:127274	LIGASE I, DNA, ATP-DEPENDENT; LIG1
	LOH19CR1	GDB:9837482	ANEMIA, CONGENITAL HYPOPLASTIC, OF BLACKFAN AND DIAMOND
30	LYL1	GDB:120158	LEUKEMIA, LYMPHOID, 1; LYL1
	MAN2B1	GDB:119376	MANNOSIDOSIS, ALPHA B, LYSOSOMAL
	MCOLN1	GDB:10013974	MUCOLIPIDOSIS IV
35	MDRV	GDB:6306714	MUSCULAR DYSTROPHY, AUTOSOMAL DOMINANT, WITH RIMMED VACUOLES; MDRV

Gene	GDB Accession ID	OMIM Link
MLLT1	GDB:136791	MYELOID/LYMPHOID OR MIXED LINEAGE LEUKEMIA, TRANSLOCATED TO, 1; MLLT1
5 NOTCH3	GDB:361163	DEMENTIA, HEREDITARY MULTI-INFARCT TYPE NOTCH, DROSOPHILA, HOMOLOG OF, 3; NOTCH3
10 NPHS1	GDB:342105	NEPHROSIS 1, CONGENITAL, FINNISH TYPE; NPHS1
15 OFC3	GDB:128060	OROFACIAL CLEFT-3; OFC3
OPA3	GDB:9954590	OPTIC ATROPHY, INFANTILE, WITH CHOREA AND SPASTIC PARAPLEGIA
PEPD	GDB:120273	PEPTIDASE D; PEPD
15 PRPF31	GDB:333911	RETINITIS PIGMENTOSA 11; RP11
PRTN3	GDB:126876	PROTEINASE 3; PRTN3; PR3
PRX	GDB:11501256	HYPERTROPHIC NEUROPATHY OF DEJERINE-SOTTAS
20 PSG1	GDB:120321	PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN 1; PSG1
PVR	GDB:120324	POLIOVIRUS SUSCEPTIBILITY, OR SENSITIVITY; PVS
25 RYR1	GDB:120359	CENTRAL CORE DISEASE OF MUSCLE HYPERTERMIA OF ANESTHESIA RYANODINE RECEPTOR-1; RYR1
SLC5A5	GDB:5892184	SOLUTE CARRIER FAMILY 5, MEMBER 5; SLC5A5
SLC7A9	GDB:9958852	CYSTINURIA, TYPE III; CSNU3
30 STK11	GDB:9732383	PEUTZ-JEGHERS SYNDROME SERINE/THREONINE PROTEIN KINASE 11; STK11
TBXA2R	GDB:127517	THROMBOXANE A2 RECEPTOR, PLATELET; TBXA2R
TGFB1	GDB:120729	ENGELMANN DISEASE TRANSFORMING GROWTH FACTOR, BETA-1; TGFB1
35 TNNI3	GDB:125309	TROPONIN I, CARDIAC; TNNI3

Gene	GDB Accession ID	OMIM Link
TYROBP	GDB:9954457	POLYCYSTIC LIPOMEMBRANOUS OSTEODYSPLOASIA WITH SCLEROSING LEUKOENCEPHALOPATHY

5 Table 21: Genes, Locations and Genetic Disorders on Chromosome 20

Gene	GDB Accession ID	OMIM Link
ADA	GDB:119649	ADENOSINE DEAMINASE; ADA
AHCY	GDB:118983	S-ADENOSYLHOMOCYSTEINE HYDROLASE; AHCY
AVP	GDB:119009	DIABETES INSIPIDUS, NEUROHYPOPHYSIAL TYPE ARGININE VASOPRESSIN; AVP
CDAN2	GDB:9823270	DYSERYTHROPOIETIC ANEMIA, CONGENITAL, TYPE II
CDMP1	GDB:438940	CHONDRODYSPLASIA, GREBE TYPE CARTILAGE-DERIVED MORPHOGENETIC PROTEIN 1
CHED1	GDB:3837719	CORNEAL DYSTROPHY, CONGENITAL ENDOTHELIAL; CHED
CHRNA4	GDB:128169	CHOLINERGIC RECEPTOR, NEURONAL NICOTINIC, ALPHA POLYPEPTIDE 4; CHRNA4 EPILEPSY, BENIGN NEONATAL; EBN1
CST3	GDB:119817	AMYLOIDOSIS VI
EDN3	GDB:119862	ENDOTHELIN-3; EDN3 WAARDENBURG-SHAH SYNDROME
EEGV1	GDB:127525	ELECTROENCEPHALOGRAM, LOW-VOLTAGE
FTLL1	GDB:119235	FERRITIN LIGHT CHAIN; FTL
GNAS	GDB:120628	GUANINE NUCLEOTIDE-BINDING PROTEIN, ALPHA-STIMULATING POLYPEPTIDE;

Gene	GDB Accession ID	OMIM Link
GSS	GDB:637022	GLUTATHIONE SYNTHETASE DEFICIENCY OF ERYTHROCYTES, HEMOLYTIC ANEMIA PYROGLUTAMICACIDURIA HNF4AGDB:393281DIABETES MELLITUS, AUTOSOMAL DOMINANT TRANSCRIPTION FACTOR 14, HEPATIC NUCLEAR FACTOR; TCF14
JAG1	GDB:6175920	CHOLESTASIS WITH PERIPHERAL PULMONARY STENOSIS JAGGED 1; JAG1
KCNQ2	GDB:9787229	EPILEPSY, BENIGN NEONATAL; EBN1 POTASSIUM CHANNEL, VOLTAGE-GATED, SUBFAMILY Q, MEMBER 2
MKKS	GDB:9860197	HYDROMETROCOLPOS SYNDROME
NBIA1	GDB:4252819	HALLERVORDEN-SPATZ DISEASE
PCK1	GDB:125349	PHOSPHOENOLPYRUVATE CARBOXYKINASE 1, SOLUBLE; PCK1
PI3	GDB:203940	PROTEINASE INHIBITOR 3; PI3
PPGB	GDB:119507	NEURAMINIDASE DEFICIENCY WITH BETA-GALACTOSIDASE DEFICIENCY
PPMD	GDB:702144	CORNEAL DYSTROPHY, HEREDITARY POLYMORPHOUS POSTERIOR; PPCD
PRNP	GDB:120720	GERSTMANN-STRAUSSLER DISEASE; GSD PRION PROTEIN; PRNP
THBD	GDB:119613	THROMBOMODULIN; THBD
TOP1	GDB:120444	TOPOISOMERASE (DNA) I; TOP1

Table 22: Genes, Locations and Genetic Disorders on Chromosome 21

Gene	GDB Accession ID	OMIM Link
AIRE	GDB:567198	AUTOIMMUNE POLYENDOCRINOPATHY-CANDIDIASIS-ECTODERMAL DYSTROPHY; APECED
APP	GDB:119692	ALZHEIMER DISEASE; AD AMYLOID BETA A4 PRECURSOR PROTEIN; APP

Gene	GDB Accession ID	OMIM Link
CBS	GDB:119754	HOMOCYSTINURIA
COL6A1	GDB:119065	COLLAGEN, TYPE VI, ALPHA-1 CHAIN; COL6A1 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
COL6A2	GDB:119793	COLLAGEN, TYPE VI, ALPHA-2 CHAIN; COL6A2 MYOPATHY, BENIGN CONGENITAL, WITH CONTRACTURES
CSTB	GDB:5215249	MYOCLONUS EPILEPSY OF UNVERRICHT AND LUNDBORG CYSTATIN B; CSTB
DCR	GDB:125354	TRISOMY 21
DSCR1	GDB:731000	TRISOMY 21
FPDMM	GDB:9954610	CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 2; CBFA2 PLATELET DISORDER, FAMILIAL, WITH ASSOCIATED MYELOID MALIGNANCY
HLCs	GDB:392648	MULTIPLE CARBOXYLASE DEFICIENCY, BIOTIN-RESPONSIVE; MCD
HPE1	GDB:136065	HOLOPROSENCEPHALY, FAMILIAL ALOBAR
ITGB2	GDB:120574	INTEGRIN BETA-2; ITGB2
KCNE1	GDB:127909	POTASSIUM VOLTAGE-GATED CHANNEL, ISK-RELATED SUBFAMILY, MEMBER 1;
KNO	GDB:4073044	KNOBLOCH SYNDROME; KNO
PRSS7	GDB:384083	ENTEROKINASE DEFICIENCY
RUNX1	GDB:128313	CORE-BINDING FACTOR, RUNT DOMAIN, ALPHA SUBUNIT 2; CBFA2 PLATELET DISORDER, FAMILIAL, WITH ASSOCIATED MYELOID MALIGNANCY

Gene	GDB Accession ID	OMIM Link
SOD1	GDB:119596	AMYOTROPHIC LATERAL SCLEROSIS SUPEROXIDE DISMUTASE-1; SOD1 MUSCULAR ATROPHY, PROGRESSIVE, WITH AMYOTROPHIC LATERAL SCLEROSIS
TAM	GDB:9958709	MYELOPROLIFERATIVE SYNDROME, TRANSIENT

Table 23: Genes, Locations and Genetic Disorders on Chromosome 22

Gene	GDB Accession ID	OMIM Link
ADSL	GDB:119655	ADENYLOSUCCINATE LYASE; ADSL
ARSA	GDB:119007	METACHROMATIC LEUKODYSTROPHY, LATE-INFANTILE
BCR	GDB:120562	BREAKPOINT CLUSTER REGION; BCR
CECR	GDB:119772	CAT EYE SYNDROME; CES
CHEK2	GDB:9958730	LI-FRAUMENI SYNDROME; LFS OSTEOGENIC SARCOMA
COMT	GDB:119795	CATECHOL-O-METHYLTRANSFERASE ; COMT
CRYBB2	GDB:119075	CRYSTALLIN, BETA B2; CRYBB2 CATARACT, CONGENITAL, CERULEAN TYPE, 2; CCA2
CSF2RB	GDB:126838	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR, BETA SUBUNIT;
CTHM	GDB:439247	HEART MALFORMATIONS; CTHM
CYP2D6	GDB:132127	CYTOCHROME P450, SUBFAMILY IID; CYP2D
CYP2D@	GDB:119832	CYTOCHROME P450, SUBFAMILY IID; CYP2D
DGCR	GDB:119843	DIGEORGE SYNDROME; DGS
DIA1	GDB:119848	METHEMOGLOBINEMIA DUE TO DEFICIENCY OF METHEMOGLOBIN REDUCTASE
EWSR1	GDB:135984	EWING SARCOMA; EWS
GGT1	GDB:120623	GLUTATHIONURIA

Gene	GDB Accession ID	OMIM Link
MGCR	GDB:120180	MENINGIOMA; MGM
MN1	GDB:580528	MENINGIOMA; MGM
5	NAGA	ALPHA-GALACTOSIDASE B; GALB
	NF2	NEUROFIBROMATOSIS, TYPE II; NF2
	OGS2	HYPERTELORISM WITH ESOPHAGEAL ABNORMALITY AND HYPOSPADIAS
	PDGFB	V-SIS PLATELET-DERIVED GROWTH FACTOR BETA POLYPEPTIDE; PDGFB
	PPARA	PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR, ALPHA; PPARA
10	PRODH	HYPERPROLINEMIA, TYPE I
	SCO2	CYTOCHROME c OXIDASE DEFICIENCY
	SCZD4	SCHIZOPHRENIA DISORDER-4; SCZD4
	SERPIND1	HEPARIN COFACTOR II; HCF2
	SLC5A1	SOLUTE CARRIER FAMILY 5, MEMBER 1; SLC5A1
15	SOX10	SRY-BOX 10; SOX10
	TCN2	TRANSCOBALAMIN II DEFICIENCY
	TIMP3	TISSUE INHIBITOR OF METALLOPROTEINASE-3; TIMP3
	VCF	VELOCARDIOFACIAL SYNDROME

Table 24: Genes, Locations and Genetic Disorders on Chromosome X

Gene	GDB Accession ID	OMIM Link
30	ABCD1	ADRENOLEUKODYSTROPHY; ALD
	ACTL1	ACTIN-LIKE SEQUENCE-1; ACTL1
	ADFN	ALBINISM-DEAFNESS SYNDROME; ADFN; ALDS
	AGMX2	AGAMMAGLOBULINEMIA, X-LINKED, TYPE 2; AGMX2; XLA2

	Gene	GDB Accession ID	OMIM Link
	AHDS	GDB:125899	MENTAL RETARDATION, X-LINKED, WITH HYPOTONIA
5	AIC	GDB:118986	CORPUS CALLOSUM, AGENESIS OF, WITH CHORIORETINAL ABNORMALITY
	AIED	GDB:119663	ALBINISM, OCULAR, TYPE 2; OA2
	AIH3	GDB:131443	AMELOGENESIS IMPERFECTA-3, HYPOPLASTIC TYPE; AIH3
10	ALAS2	GDB:119666	ANEMIA, HYPOCHROMIC
	AMCD	GDB:5584286	ARTHROGRYPOSIS MULTIPLEX CONGENITA, DISTAL
	AMELX	GDB:119675	AMELOGENESIS IMPERFECTA-1, HYPOPLASTIC TYPE; AIH1
15	ANOP1	GDB:128454	CLINICAL; ANOP1
	AR	GDB:120556	ANDROGEN INSENSITIVITY SYNDROME; AIS ANDROGEN RECEPTOR; AR
	ARAF1	GDB:119004	V-RAF MURINE SARCOMA 3611 VIRAL ONCOGENE HOMOLOG 1; ARAF1
20	ARSC2	GDB:119702	ARYLSULFATASE C, f FORM; ARSC2
	ARSE	GDB:555743	CHONDRODYSPLASIA PUNCTATA 1, X-LINKED RECESSIVE; CDPX1
	ARTS	GDB:9954651	FATAL X-LINKED, WITH DEAFNESS AND LOSS OF VISION
25	ASAT	GDB:9954649	SIDEROBLASTIC, AND SPINOCEREBELLAR ATAXIA; ASAT
	ASSP5	GDB:119019	CITRULLINEMIA
	ATP7A	GDB:119395	ATPase, Cu(2+)-TRANSPORTING, ALPHA POLYPEPTIDE; ATP7A MENKES SYNDROME
30	ATRX	GDB:136052	ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, X-LINKED; ATRX ALPHA-THALASSEMIA/MENTAL RETARDATION SYNDROME, NONDELETION TYPE
	AVPR2	GDB:131475	DIABETES INSIPIDUS, NEPHROGENIC
35	BFLS	GDB:120566	BORJESON SYNDROME; BORJ

Gene	GDB Accession ID	OMIM Link
BGN	GDB:119727	BIGLYCAN; BGN
5 BTK	GDB:120542	BRUTON AGAMMAGLOBULINEMIA TYROSINE KINASE; BTK
BZX	GDB:5205912	BAZEX SYNDROME; BZX
C1HR	GDB:119040	TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR 2A; TAF2A
10 CACNA1F	GDB:6053864	NIGHTBLINDNESS, CONGENITAL STATIONARY, X-LINKED, TYPE 2; CSNB2 CALCIUM CHANNEL, VOLTAGE-DEPENDENT, ALPHA 1F SUBUNIT; CACNA1F
CALB3	GDB:133780	CALBINDIN 3; CALB3
15 CBBM	GDB:9958963	COLORBLINDNESS, BLUE-MONO-CONE-MONOCHROMATIC TYPE; CBBM
CCT	GDB:119756	CATARACT, CONGENITAL TOTAL, WITH POSTERIOR SUTURAL OPACITIES IN HETEROZYGOTES;
20 CDR1	GDB:119053	CEREBELLAR DEGENERATION-RELATED AUTOANTIGEN-1; CDR1; CDR34
CFNS	GDB:9579470	CRANIOFRONTONASAL SYNDROME; CFNS
CGF1	GDB:6275867	COGNITION
25 CHM	GDB:120400	CHOROIDEREMIA; CHM
CHR39C	GDB:119779	CHOLESTEROL REPRESSIBLE PROTEIN 39C; CHR39C
CIDX	GDB:127736	SEVERE COMBINED IMMUNODEFICIENCY DISEASE, X-LINKED, 2; SCIDX2
30 CLA2	GDB:119782	CEREBELLAR ATAXIA, X-LINKED; CLA2
CLCN5	GDB:270667	CHLORIDE CHANNEL 5; CLCN5 FANCONI SYNDROME, RENAL, WITH NEPHROCALCINOSIS AND RENAL STONES NEPHROLITHIASIS, X-LINKED RECESSIVE, WITH RENAL FAILURE; XRN
35 CLS	GDB:119784	RIBOSOMAL PROTEIN S6 KINASE, 90 KD, POLYPEPTIDE 3; RPS6KA3 COFFIN-LOWRY SYNDROME; CLS

Gene	GDB Accession ID	OMIM Link
CMTX2	GDB:128311	CHARCOT-MARIE-TOOTH NEUROPATHY, X-LINKED RECESSIVE, 2; CMTX2
5 CMTX3	GDB:128151	CHARCOT-MARIE-TOOTH NEUROPATHY, X-LINKED RECESSIVE, 3; CMTX3
CND	GDB:9954627	DERMOIDS OF CORNEA; CND
10 COD1	GDB:119787	CONE DYSTROPHY, X-LINKED, 1; COD1
COD2	GDB:6520166	CONE DYSTROPHY, X-LINKED, 2; COD2
15 COL4A5	GDB:120596	COLLAGEN, TYPE IV, ALPHA-5 CHAIN; COL4A5 LEIOMYOMATOSIS, ESOPHAGEAL AND VULVAL, WITH NEPHROPATHY
COL4A6	GDB:222775	COLLAGEN, TYPE IV, ALPHA-6 CHAIN; COL4A6 LEIOMYOMATOSIS, ESOPHAGEAL AND VULVAL, WITH NEPHROPATHY
CPX	GDB:120598	CLEFT PALATE, X-LINKED; CPX
18 CVD1	GDB:9954659	CARDIAC VALVULAR DYSPLASIA, X-LINKED
20 CYBB	GDB:120513	GRANULOMATOUS DISEASE, CHRONIC; CGD
DCX	GDB:9823272	LISSENCEPHALY, X-LINKED
25 DFN2	GDB:119091	DEAFNESS, X-LINKED 2, PERCEPTIVE CONGENITAL; DFN2
DFN4	GDB:433255	DEAFNESS, X-LINKED 4, CONGENITAL SENSORINEURAL; DFN4
28 DFN6	GDB:1320698	DEAFNESS, X-LINKED, 6, PROGRESSIVE; DFN6
DHOF	GDB:119847	FOCAL DERMAL HYPOPLASIA; DHOF
30 DIAPH2	GDB:9835484	DIAPHANOUS, DROSOPHILA, HOMOLOG OF, 2 DKC1GDB:119096 DYSKERATOSIS CONGENITA; DKC
DMD	GDB:119850	MUSCULAR DYSTROPHY, PSEUDOHYPERTROPHIC PROGRESSIVE, DUCHENNE AND BECKER
35 DSS	GDB:433750	DOSAGE-SENSITIVE SEX REVERSAL; DSS

	Gene	GDB Accession ID	OMIM Link
	DYT3	GDB:118789	TORSION DYSTONIA-3, X-LINKED TYPE; DYT3
5	EBM	GDB:119102	BULLOUS DYSTROPHY, HEREDITARY MACULAR TYPE
	EBP	GDB:125212	CHONDRODYSPLASIA PUNCTATA, X-LINKED DOMINANT; CDPX2; CDPXD; CPXD
10	ED1	GDB:119859	ECTODERMAL DYSPLASIA, ANHIDROTIC; EDA
	ELK1	GDB:119867	ELK1, MEMBER OF ETS ONCOGENE FAMILY; ELK1
15	EMD	GDB:119108	MUSCULAR DYSTROPHY, TARDIVE, DREIFUSS-EMERY TYPE, WITH CONTRACTURES
	EVR2	GDB:136068	EXUDATIVE VITREORETINOPATHY, FAMILIAL, X-LINKED RECESSIVE; EVR2
	F8C	GDB:119124	HEMOPHILIA A
	F9	GDB:119900	HEMOPHILIA B; HEMB
20	FCP1	GDB:347490	F-CELL PRODUCTION, X-LINKED; FCPX
	FDPSL5	GDB:119922	SYNTHETASE-5; FPSL5
	FGD1	GDB:119131	SYNDROME FACIOGENITAL DYSPLASIA; FGDY
	FGS1	GDB:9836950	FG SYNDROME
25	FMR1	GDB:129038	FRAGILE SITE MENTAL RETARDATION-1; FMR1
	FMR2	GDB:141566	FRAGILE SITE, FOLIC ACID TYPE, RARE, FRA(X)(q28); FRAXE
30	G6PD	GDB:120621	GLUCOSE-6-PHOSPHATE DEHYDROGENASE; G6PD
	GABRA3	GDB:119968	GAMMA-AMINOBUTYRIC ACID RECEPTOR, ALPHA-3; GABRA3
	GATA1	GDB:125373	GATA-BINDING PROTEIN 1; GATA1
35	GDI1	GDB:1347097	GDP DISSOCIATION INHIBITOR 1; GDI1 MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 3; MRX3

Gene	GDB Accession ID	OMIM Link
GDXY	GDB:9954629	DYSGENESIS, XY FEMALE TYPE; GDXY
5 GJB1	GDB:125246	CHARCOT-MARIE-TOOTH PERONEAL MUSCULAR ATROPHY, X-LINKED; CMTX1 GAP JUNCTION PROTEIN, BETA-1, 32 KD; GJB1
GK	GDB:119271	HYPERGLYCEROLEMIA
GLA	GDB:119272	ANGIOKERATOMA, DIFFUSE
10 GPC3	GDB:3770726	GLYCAN-3; GPC3 SIMPSON DYSMORPHIA SYNDROME; SDYS
GRPR	GDB:128035	GASTRIN-RELEASING PEPTIDE RECEPTOR; GRPR
GTD	GDB:9954635	GONADOTROPIN DEFICIENCY; GTD
15 GUST	GDB:9954655	MENTAL RETARDATION WITH OPTIC ATROPHY, DEAFNESS, AND SEIZURES
HMS1	GDB:251827	1; HMS1
20 HPRT1	GDB:119317	HYPOXANTHINE GUANINE PHOSPHORIBOSYLTRANSFERASE 1; HPRT1
HPT	GDB:119322	HYPOPARATHYROIDISM, X-LINKED; HYPX
HTC2	GDB:700980	HYPERTRICHOSIS, CONGENITAL GENERALIZED; CGH; HCG
25 HTR2C	GDB:378202	5-@HYDROXYTRYPTAMINE RECEPTOR 2C; HTR2C
HYR	GDB:9954625	REGULATOR; HYR
IDS	GDB:120521	MUCOPOLYSACCHARIDOSIS TYPE II
IHG1	GDB:119343	HYPOPLASIA OF, WITH GLAUCOMA; IHG
30 IL2RG	GDB:134807	INTERLEUKIN-2 RECEPTOR, GAMMA; IL2RG SEVERE COMBINED IMMUNODEFICIENCY DISEASE, X-LINKED, 2; SCIDX2
INDX	GDB:9954657	IMMUNONEUROLOGIC DISORDER, X-LINKED
35 IP1	GDB:120105	INCONTINENTIA PIGMENTI, TYPE I; IP1

Gene	GDB Accession ID	OMIM Link
IP2	GDB:120106	INCONTINENTIA PIGMENTI, TYPE II; IP2
5 JMS	GDB:204055	MENTAL RETARDATION, X-LINKED, WITH GROWTH RETARDATION, DEAFNESS, AND
KAL1	GDB:120116	KALLMANN SYNDROME 1; KAL1
KFSD	GDB:128174	KERATOSIS FOLLICULARIS SPINULOSA DECALVANS CUM OPHIASI; KFSD
10 L1CAM	GDB:120133	CLASPED THUMB AND MENTAL RETARDATION L1 CELL ADHESION MOLECULE; L1CAM
LAMP2	GDB:125376	LYSOSOME-ASSOCIATED MEMBRANE PROTEIN B; LAMP2; LAMPB
15 MAA	GDB:119372	MICROPHTHALMIA OR ANOPHTHALMOS, WITH ASSOCIATED ANOMALIES; MAA
MAFD2	GDB:119373	PSYCHOSIS, X-LINKED
MAOA	GDB:120164	MONOAMINE OXIDASE A; MAOA
MAOB	GDB:119377	MONOAMINE OXIDASE B; MAOB
20 MCF2	GDB:120168	MCF.2 CELL LINE DERIVED TRANSFORMING SEQUENCE; MCF2
MCS	GDB:128370	MENTAL RETARDATION, X-LINKED, SYNDROMIC-4, WITH CONGENITAL CONTRACTURES
25 MEAX	GDB:119383	X-LINKED, WITH EXCESSIVE AUTOPHAGY; XMEA; MEAX
MECP2	GDB:3851454	SYNDROME; RTT
MF4	GDB:119386	METACARPAL 4-5 FUSION; MF4
MGC1	GDB:120179	MEGALOCORNEA; MGC1; MGCN
30 MIC5	GDB:120526	SURFACE ANTIGEN, X-LINKED; SAX
MD1	GDB:9772232	OPITZ SYNDROME
MLLT7	GDB:392309	MYELOID/LYMPHOID OR MIXED-LINEAGE LEUKEMIA, TRANSLOCATED TO, 7; MLLT7
35 MLS	GDB:262123	MICROPHTHALMIA WITH LINEAR SKIN DEFECTS; MLS

Gene	GDB Accession ID	OMIM Link
5	MRSD	MENTAL RETARDATION, SKELETAL DYSPLASIA, AND ABDUCENS PALSY; MRSD
	MRX14	RETARDATION, X-LINKED 14; MRX14
	MRX1	MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 1; MRX1
	MRX20	MENTAL RETARDATION, X-LINKED 20; MRX20
10	MRX2	RETARDATION, X-LINKED NONSPECIFIC, TYPE 2; MRX2
	MRX3	GDP DISSOCIATION INHIBITOR 1; GDI1 MENTAL RETARDATION, X-LINKED NONSPECIFIC, TYPE 3; MRX3
	MRX40	MENTAL RETARDATION, X-LINKED, WITH HYPOTONIA
	MRXA	MENTAL RETARDATION, X-LINKED NONSPECIFIC, WITH APHASIA; MRXA
15	MSD	SYNDROME
	MTM1	MYOTUBULAR MYOPATHY 1; MTM1
	MYCL2	MYCL-RELATED PROCESSED GENE; MYCL2
	MYP1	MYOPIA, X-LINKED; MYP1
20	NDP	NORRIE DISEASE; NDP
	NHS	CATARACT-DENTAL SYNDROME
	NPHL1	NEPHROLITHIASIS, X-LINKED RECESSIVE, WITH RENAL FAILURE; XRN
	NR0B1	ADRENAL HYPOPLASIA, CONGENITAL; AHC
25	NSX	SYNDROME; NSX
	NYS1	NYSTAGMUS, X-LINKED; NYS
	NYX	NIGHTBLINDNESS, CONGENITAL STATIONARY, WITH MYOPIA; CSNB1
	OA1	ALBINISM, OCULAR, TYPE 1; OA1
30		
35		

	Gene	GDB Accession ID	OMIM Link
	OASD	GDB:138457	OCULAR, WITH LATE-ONSET SENSORINEURAL DEAFNESS; OASD
5	OCRL	GDB:119461	LOWE OCULOCEREBRORENAL SYNDROME; OCRL
	ODT1	GDB:125360	TEETH, ABSENCE OF
	OFD1	GDB:120248	OROFACIODIGITAL SYNDROME 1; OFD1
10	OPA2	GDB:125358	OPTIC ATROPHY 2; OPA2
	OPD1	GDB:120249	OTOPALATODIGITAL SYNDROME
	OPEM	GDB:119467	OPHTHALMOPLEGIA, EXTERNAL, AND MYOPIA; OPEM
	OPN1LW	GDB:120724	COLORBLINDNESS, PARTIAL, PROTAN SERIES; CBP
15	OPN1MW	GDB:120622	COLORBLINDNESS, PARTIAL, DEUTAN SERIES; CBD; DCB
	OTC	GDB:119468	ORNITHINE TRANSCARBAMYLASE DEFICIENCY, HYPERAMMONEMIA DUE TO; OTC
20	P3	GDB:9954667	PROTEIN P3
	PDHA1	GDB:118895	PYRUVATE DEHYDROGENASE COMPLEX, E1-ALPHA POLYPEPTIDE-1; PDHA1
	PDR	GDB:203409	AMYLOIDOSIS, FAMILIAL CUTANEOUS
	PFC	GDB:120275	PROPERDIN DEFICIENCY, X-LINKED
25	PFKFB1	GDB:125375	6-@PHOSPHOFRUCTO-2-KINASE; PFKFB1
	PGK1	GDB:120282	PHOSPHOGLYCERATE KINASE 1; PGK1
	PGK1P1	GDB:120283	PHOSPHOGLYCERATE KINASE 1; PGK1
30	PGS	GDB:128372	DANDY-WALKER MALFORMATION WITH MENTAL RETARDATION, BASAL GANGLIA DISEASE,
	PHEX	GDB:120520	HYPOPHOSPHATEMIA, VITAMIN D-RESISTANT RICKETS; HYP
	PHKA1	GDB:120285	PHOSPHORYLASE KINASE, ALPHA 1 SUBUNIT (MUSCLE); PHKA1
35	PHKA2	GDB:127279	GLYCOGEN STORAGE DISEASE VIII

	Gene	GDB Accession ID	OMIM Link
5	PHP	GDB:119494	PANHYPOPITUITARISM; PHP
	PIGA	GDB:138138	PHOSPHATIDYLINOSITOL GLYCAN, CLASS A; PIGA
	PLP1	GDB:120302	PROTEOLIPID PROTEIN, MYELIN; PLP
	POF1	GDB:120716	PREMATURE OVARIAN FAILURE 1; POF1
	POLA	GDB:120304	POLYMERASE, DNA, ALPHA; POLA
10	POU3F4	GDB:351386	DEAFNESS, CONDUCTIVE, WITH STAPES FIXATION
	PPMX	GDB:9954669	RETARDATION WITH PSYCHOSIS, PYRAMIDAL SIGNS, AND MACROORCHIDISM
	PRD	GDB:371323	DYSPLASIA, PRIMARY
	PRPS1	GDB:120318	PHOSPHORIBOSYL PYROPHOSPHATE SYNTHETASE-I; PRPS1
	PRPS2	GDB:120320	PHOSPHORIBOSYL PYROPHOSPHATE SYNTHETASE-II; PRPS2
15	PRS	GDB:128368	MENTAL RETARDATION, X-LINKED, SYNDROMIC-2, WITH DYSMORPHISM AND CEREBRAL
	PRTS	GDB:128367	PARTINGTON X-LINKED MENTAL RETARDATION SYNDROME; PRTS
	PSF2	GDB:119519	TRANSPORTER 2, ABC; TAP2
	RENBP	GDB:133792	RENIN-BINDING PROTEIN; RENBP
	RENS1	GDB:9806348	MENTAL RETARDATION, X-LINKED, RENPENNING TYPE
20	RP2	GDB:120353	RETINITIS PIGMENTOSA-2; RP2
	RP6	GDB:125381	PIGMENTOSA-6; RP6
	RPGR	GDB:118736	RETINITIS PIGMENTOSA-3; RP3
	RPS4X	GDB:128115	RIBOSOMAL PROTEIN S4, X-LINKED; RPS4X
	RPS6KA3	GDB:365648	RIBOSOMAL PROTEIN S6 KINASE, 90 KD, POLYPEPTIDE 3; RPS6KA3
25	RS1	GDB:119581	RETINOSCHISIS; RS

Gene	GDB Accession ID	OMIM Link
S11	GDB:120361	ANTIGEN, X-LINKED, SECOND; SAX2
5 SDYS	GDB:119590	GLYCAN-3; GPC3 SIMPSON DYSMORPHIA SYNDROME; SDYS
	GDB:120372	SPONDYLOEPIPHYSEAL DYSPLASIA, LATE; SEDL
	GDB:120399	THYROXINE-BINDING GLOBULIN OF SERUM; TBG
10 SH2D1A	GDB:120701	IMMUNODEFICIENCY, X-LINKED PROGRESSIVE COMBINED VARIABLE
	GDB:226635	SPLIT-HAND/SPLIT-FOOT ANOMALY, X-LINKED
	GDB:6118451	SHORT STATURE; SS
15 SLC25A5	GDB:125190	ADENINE NUCLEOTIDE TRANSLOCATOR 2; ANT2
	GDB:9954643	SPINAL MUSCULAR ATROPHY, X-LINKED LETHAL INFANTILE
	GDB:3811398	RETINITIS PIGMENTOSA-3; RP3
20 SRS	GDB:136337	MENTAL RETARDATION, X-LINKED, SNYDER-ROBINSON TYPE
	GDB:120393	ICHTHYOSIS, X-LINKED
	GDB:119606	SYNAPSIN I; SYN1
25 SYP	GDB:125295	SYNAPTOPHYSIN; SYP
	GDB:120573	TATA BOX BINDING PROTEIN (TBP)-ASSOCIATED FACTOR 2A; TAF2A
	GDB:120609	CARDIOMYOPATHY, DILATED 3A; CMD3A ENDOCARDIAL FIBROELASTOSIS-2; EFE2
30 TBX22	GDB:10796448	CLEFT PALATE, X-LINKED; CPX
	GDB:119610	MALE PSEUDOHERMAPHRODITISM: DEFICIENCY OF TESTICULAR 17,20-DESMOLASE;
	GDB:125870	TRANSCRIPTION FACTOR FOR IMMUNOGLOBULIN HEAVY-CHAIN ENHANCER-3; TFE3
35 THAS	GDB:128158	THORACOABDOMINAL SYNDROME; TAS

Gene	GDB Accession ID	OMIM Link
THC	GDB:125361	THROMBOCYTOPENIA, X-LINKED; THC; XLT
5 TIMM8A	GDB:119090	DEAFNESS 1, PROGRESSIVE; DFN1
TIMP1	GDB:119615	TISSUE INHIBITOR OF METALLOPROTEINASE-1; TIMP1
10 TKCR	GDB:119616	TORTICOLLIS, KELOIDS, CRYPTORCHIDISM, AND RENAL DYSPLASIA; TKC
TNFSF5	GDB:120632	IMMUNODEFICIENCY WITH INCREASED IgM
UBE1	GDB:118954	UBIQUITIN-ACTIVATING ENZYME 1; UBE1
UBE2A	GDB:131647	UBIQUITIN-CONJUGATING ENZYME E2A; UBE2A
15 WAS	GDB:120736	WISKOTT-ALDRICH SYNDROME; WAS
WSN	GDB:125864	PARKINSONISM, EARLY-ONSET, WITH MENTAL RETARDATION
20 WTS	GDB:128373	MENTAL RETARDATION, X-LINKED, SYNDROMIC-6, WITH GYNECOMASTIA AND OBESITY;
WWS	GDB:120497	WIEACKER SYNDROME
XIC	GDB:120498	X-INACTIVATION-SPECIFIC TRANSCRIPT; XIST
25 XIST	GDB:126428	X-INACTIVATION-SPECIFIC TRANSCRIPT; XIST
XK	GDB:120499	Xk LOCUS
XM	GDB:119634	XM SYSTEM
XS	GDB:119636	LUTHERAN SUPPRESSOR, X-LINKED; XS; LUXS
30 ZFX	GDB:120502	ZINC FINGER PROTEIN, X-LINKED; ZFX
ZIC3	GDB:249141	HETEROTAXY, X-LINKED VISCERAL; HTX1
35 ZNF261	GDB:9785766	MENTAL RETARDATION, X-LINKED; DDX6673E
ZNF41	GDB:125865	ZINC FINGER PROTEIN-41; ZNF41

Gene	GDB Accession ID	OMIM Link
ZNF6	GDB:120508	ZINC FINGER PROTEIN-6; ZNF6

Table 25: Genes, Locations and Genetic Disorders on Chromosome Y

Gene	GDB Accession ID	OMIM Link
AMELY	GDB:119676	AMELOGENIN, Y-CHROMOSOMAL; AMELY
ASSP6	GDB:119020	CITRULLINEMIA
AZF1	GDB:119027	AZOOSPERMIA FACTOR 1; AZF1
AZF2	GDB:456131	AZOOSPERMIA FACTOR 2; AZF2
DAZ	GDB:635890	DELETED IN AZOOSPERMIA; DAZ
GCY	GDB:119267	CONTROL, Y-CHROMOSOME INFLUENCED; GCY
RPS4Y	GDB:128052	RIBOSOMAL PROTEIN S4, Y-LINKED; RPS4Y
SMCY	GDB:5875390	HISTOCOMPATIBILITY Y ANTIGEN; HY; HYA
SRY	GDB:125556	SEX-DETERMINING REGION Y; SRY
ZFY	GDB:120503	ZINC FINGER PROTEIN, Y-LINKED; ZFY

Table 26: Genes, Locations and Genetic Disorders in Unknown or Multiple Locations

Gene	GDB Accession ID	OMIM Link
ABAT	GDB:581658	GAMMA-AMINOBUTYRATE TRANSAMINASE
AEZ	GDB:128360	ACRODERMATITIS ENTEROPATHICA, ZINC-DEFICIENCY TYPE; AEZ
AFA	GDB:265277	FILIFORME ADNATUM AND CLEFT PALATE
AFD1	GDB:265292	DYSOSTOSIS, TREACHER COLLINS TYPE, WITH LIMB ANOMALIES
AGS1	GDB:10795417	ENCEPHALOPATHY, FAMILIAL INFANTILE, WITH CALCIFICATION OF BASAL GANGLIA
ASAH	GDB:6837715	FARBER LIPOGRANULOMATOSIS
ASD1	GDB:6276019	ATRIAL SEPTAL DEFECT; ASD

Gene	GDB Accession ID	OMIM Link
ASMT	GDB:136259	CETYLSEROTONIN METHYLTRANSFERASE; ASMT ACETYL SEROTONIN METHYLTRANSFERASE, Y-CHROMOSOMAL; ASMTY; HIOMTY
BCH	GDB:118758	CHOREA, HEREDITARY BENIGN; BCH
CCAT	GDB:118738	CATARACT, CONGENITAL OR JUVENILE
CECR9	GDB:10796163	CAT EYE SYNDROME; CES
CEPA	GDB:581848	CONTROL, CONGENITAL FAILURE OF
CHED2	GDB:9957389	CORNEAL DYSTROPHY, CONGENITAL HEREDITARY
CLA1	GDB:119781	CEREBELLOPARENCHYMAL DISORDER III
CLA3	GDB:128453	CEREBELLOPARENCHYMAL DISORDER I; CPD I
CLN4	GDB:125229	CEROID-LIPOFUSCINOSIS, NEURONAL 4; CLN4
CPO	GDB:119070	COPROPORPHYRIA
CSF2RA	GDB:118777	COLONY STIMULATING FACTOR 2 RECEPTOR, ALPHA; CSF2RA GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR, ALPHA SUBUNIT,
CTS1	GDB:118779	CARPAL TUNNEL SYNDROME; CTS; CTS1
DF	GDB:132645	FACTOR D
DIH1	GDB:439243	DIAPHRAGMATIC
DWS	GDB:128371	SYNDROME; DWS
DYT2	GDB:118788	DYSTONIA MUSCULORUM DEFORMANS 2; DYT2
DYT4	GDB:433751	DYSTONIA MUSCULORUM DEFORMANS 4; DYT4
EBR3	GDB:118739	EPIDERMOLYSIS BULLOSA DYSTROPHICA NEUROTROPHICA
ECT	GDB:128640	CENTRALOPATHIC EPILEPSY
EEF1A1L14	GDB:1327185	PROSTATIC CARCINOMA ONCOGENE PTI-1
EYCL2	GDB:4642815	EYE COLOR-3; EYCL3

Gene	GDB Accession ID	OMIM Link
FA1	GDB:118795	FANCONI ANEMIA, COMPLEMENTATION GROUP A; FACA
5 FANCB	GDB:9864269	FANCONI PANCYTOPENIA, TYPE 2
GCSH	GDB:126842	HYPERGLYCINEMIA, ISOLATED NONKETOTIC, TYPE III; NKH3
GCSL	GDB:132139	ISOLATED NONKETOTIC, TYPE IV; NKH4
10 GDF5	GDB:433948	CARTILAGE-DERIVED MORPHOGENETIC PROTEIN 1
GIP	GDB:119985	GASTRIC INHIBITORY POLYPEPTIDE; GIP
GTS	GDB:118807	GILLES DE LA TOURETTE SYNDROME; GTS
HHG	GDB:118740	HYPERGONADOTROPIC HYPOGONADISM; HHG
15 HMI	GDB:265275	OF ITO; HMI
HOAC	GDB:118812	DEAFNESS, CONGENITAL, AUTOSOMAL RECESSIVE
HOKPP2	GDB:595535	HYPOKALEMIC PERIODIC PARALYSIS, TYPE II; HOKPP2
20 HRPT1	GDB:125252	HYPERPARATHYROIDISM, FAMILIAL PRIMARY
HSD3B3	GDB:676973	GIANT CELL HEPATITIS, NEONATAL
HTC1	GDB:265286	HYPERTRICHOSIS UNIVERSALIS CONGENITA, AMBRAS TYPE; HTC1
HV1S	GDB:9955009	HERPES VIRUS SENSITIVITY; HV1S
25 ICR1	GDB:127785	LAMELLAR, AUTOSOMAL DOMINANT FORM
ICR5	GDB:127789	ICHTHYOSIS CONGENITA, HARLEQUIN FETUS TYPE
IL3RA	GDB:128985	INTERLEUKIN-3 RECEPTOR, ALPHA; IL3RA INTERLEUKIN-3 RECEPTOR, Y-CHROMOSOMAL; IL3RA
30 KAL2	GDB:265288	KALLMANN SYNDROME 2; KAL2
KMS	GDB:118827	SYNDROME; KMS
KRT18	GDB:120127	KERATIN 18; KRT18
35 KSS	GDB:9957718	KEARNS-SAYRE SYNDROME; KSS

Gene	GDB Accession ID	OMIM Link
LCAT	GDB:119359	FISH-EYE DISEASE; FED LECITHIN:CHOLESTEROL ACYLTRANSFERASE DEFICIENCY
5 LIMM	GDB:9958161	MYOPATHY, MITOCHONDRIAL, LETHAL INFANTILE; LIMM
MANBB	GDB:125262	MANNOSIDOSIS, BETA; MANB1
MCPH2	GDB:9863035	MICROCEPHALY; MCT
10 MEB	GDB:599557	DISEASE
MELAS	GDB:9955855	MELAS SYNDROME
MIC2	GDB:120184	SURFACE ANTIGEN MIC2; MIC2; CD99 MIC2 SURFACE ANTIGEN, Y-CHROMOSOMAL; MIC2Y
15 MPFD	GDB:439372	CONGENITAL, WITH FIBER-TYPE DISPROPORTION
MS	GDB:229116	SCLEROSIS; MS
MSS	GDB:118743	MARINESCO-SJOGREN SYNDROME; MSS
MTATP6	GDB:118897	ATP SYNTHASE 6; MTATP6
20 MTCO1	GDB:118900	COMPLEX IV, CYTOCHROME c OXIDASE SUBUNIT I; MTCO1; COI
MTCO3	GDB:118902	CYTOCHROME c OXIDASE III; MTCO3
MTCYB	GDB:118906	COMPLEX III, CYTOCHROME b SUBUNIT
MTND1	GDB:118911	COMPLEX I, SUBUNIT ND1; MTND1
25 MTND2	GDB:118912	COMPLEX I, SUBUNIT ND2; MTND2
MTND4	GDB:118914	COMPLEX I, SUBUNIT ND4; MTND4
MTND5	GDB:118916	COMPLEX I, SUBUNIT ND5; MTND5
MTND6	GDB:118917	COMPLEX I, SUBUNIT ND6; MTND6
30 MTRNR1	GDB:118920	RIBOSOMAL RNA, MITOCHONDRIAL, 12S; MTRNR1
MTRNR2	GDB:118921	RIBOSOMAL RNA, MITOCHONDRIAL, 16S; MTRNR2
35 MTTE	GDB:118926	TRANSFER RNA, MITOCHONDRIAL, GLUTAMIC ACID; MTTE

Gene	GDB Accession ID	OMIM Link
MTTG	GDB:118933	TRANSFER RNA, MITOCHONDRIAL, GLYCINE; MTTG
5 MTTI	GDB:118935	TRANSFER RNA, MITOCHONDRIAL, ISOLEUCINE; MTTI
MTTK	GDB:118936	MERRF SYNDROME TRANSFER RNA, MITOCHONDRIAL, LYSINE; MTTK
10 MTTL1	GDB:118937	MERRF SYNDROME TRANSFER RNA, MITOCHONDRIAL, LEUCINE, 1; MTTL1
MTTL2	GDB:118938	TRANSFER RNA, MITOCHONDRIAL, LEUCINE, 2; MTTL2
15 MTTN	GDB:118940	TRANSFER RNA, MITOCHONDRIAL, ASPARAGINE; MTTN
MTTP	GDB:118941	TRANSFER RNA, MITOCHONDRIAL, PROLINE; MTTP
MTTS1	GDB:118944	TRANSFER RNA, MITOCHONDRIAL, SERINE, 1; MTTS1
NAMSD	GDB:681237	NEUROPATHY, MOTOR-SENSORY, TYPE II, WITH DEAFNESS AND MENTAL RETARDATION
20 NODAL	GDB:9848762	NODAL, MOUSE, HOMOLOG OF
OCD1	GDB:118846	DISORDER-1; OCD1
OPD2	GDB:131394	SYNDROME
25 PCK2	GDB:137198	PHOSPHOENOLPYRUVATE CARBOXYKINASE 2, MITOCHONDRIAL; PCK2
PCLD	GDB:433949	POLYCYSTIC LIVER DISEASE; PLD
PCOS1	GDB:1391802	STEIN-LEVENTHAL SYNDROME
PFKM	GDB:120277	GLYCOGEN STORAGE DISEASE VII
30 PKD3	GDB:127866	KIDNEY DISEASE 3, AUTOSOMAL DOMINANT; PKD3
PRCA1	GDB:342066	PROSTATE CANCER; PRCA1
PRO1	GDB:128585	
PROP1	GDB:9834318	PROPHET OF PIT1, MOUSE, HOMOLOG OF; PROP1
35 RBS	GDB:118862	ROBERTS SYNDROME; RBS

Gene	GDB Accession ID	OMIM Link
RFXAP	GDB:9475355	REGULATORY FACTOR X-ASSOCIATED PROTEIN; RFXAP
5 RP	GDB:9958158	RETINITIS PIGMENTOSA-8
SLC25A6	GDB:125184	ADELINE NUCLEOTIDE TRANSLOCATOR 3; ANT3 ADELINE NUCLEOTIDE TRANSLOCATOR 3, Y-CHROMOSOMAL; ANT3Y
10 SPG5B	GDB:250333	SPASTIC PARAPLEGIA-5B, AUTOSOMAL RECESSIVE; SPG5B
STO	GDB:439375	CEREBRAL GIGANTISM
SUOX	GDB:5584405	SULFOCYSTEINURIA
TC21	GDB:5573831	ONCOGENE TC21
15 THM	GDB:439378	FAMILIAL
TST	GDB:134043	RHODANESE; RDS
TTD	GDB:230276	TRICHOThIODYSTROPHY; TTD

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

25

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

5 1. A method for identifying a test compound that modulates premature translation termination or nonsense-mediated mRNA decay, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA: test compound complex is formed;

10 (b) separating the detectably labeled target RNA: test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA: test compound complex.

2. The method of paragraph 1 in which the target RNA molecule  
15 contains regions of 28S rRNA or analogs thereof.

3. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.

20 4. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

25 5. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.

30 6. The method of paragraph 5 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

7. The method of paragraph 5 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl<sub>2</sub>. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl<sub>2</sub>. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

8. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.

9. The structure of the test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

30

35

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that modulates premature  
5 translation termination or nonsense-mediated mRNA decay, comprising the steps of:
  - (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
  - 10 (b) separating the detectably labeled target RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds by capillary gel electrophoresis; and
  - 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex by mass spectroscopy.

20

25

30

35

ABSTRACT

The present invention relates to a method for screening and identifying test compounds that modulate premature translation termination and/or nonsense-mediated messenger ribonucleic acid ("mRNA") by interacting with a preselected target ribonucleic acid ("RNA"). In particular, the present invention relates to identifying test compounds that bind to regions of the 28S ribosomal RNA ("rRNA") and analogs thereof. Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

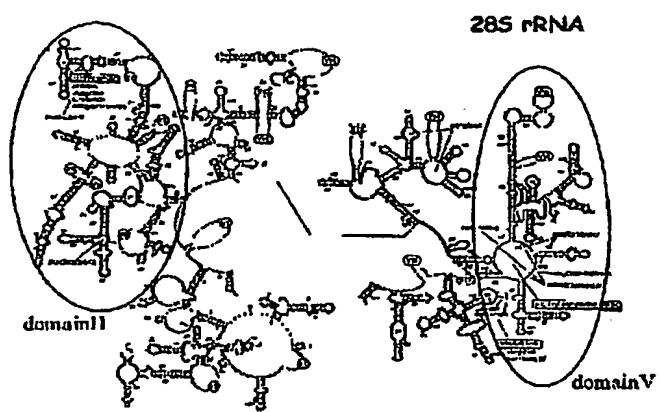
20

25

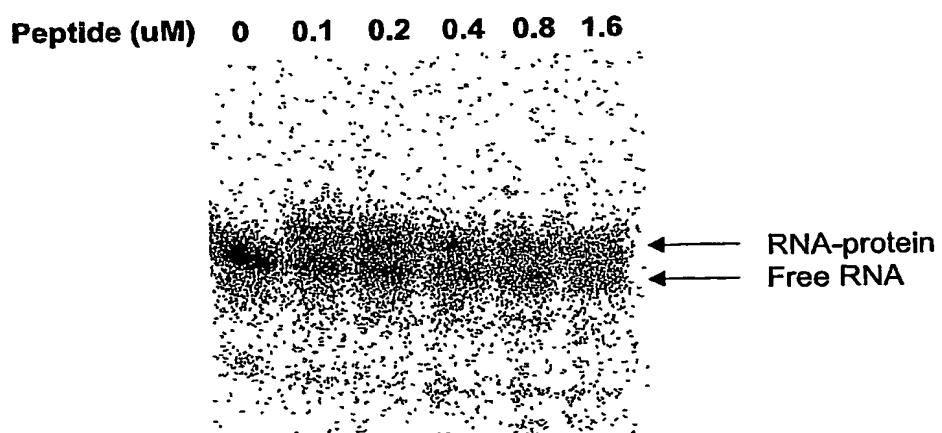
30

35

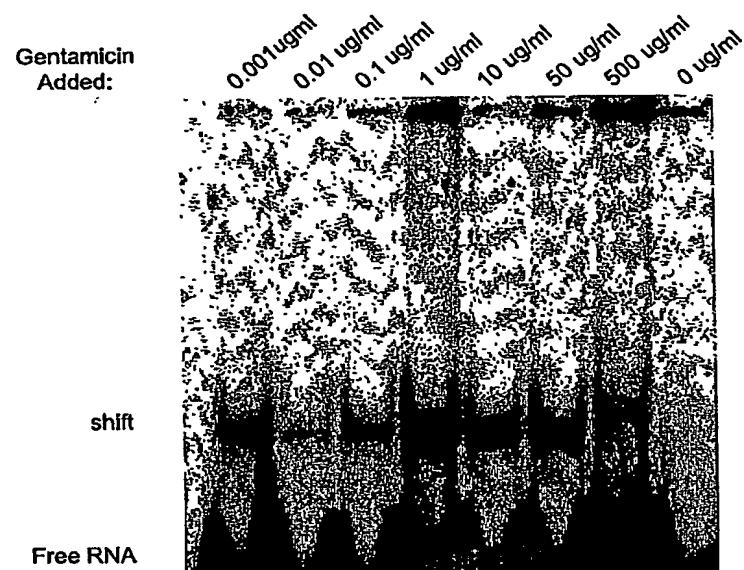
**Figure 1**  
**Sheet 1/11**  
**Attorney Docket No. 10589-021**



**Figure 2**  
**Sheet 2/11**  
**Attorney Docket No. 10589-021**

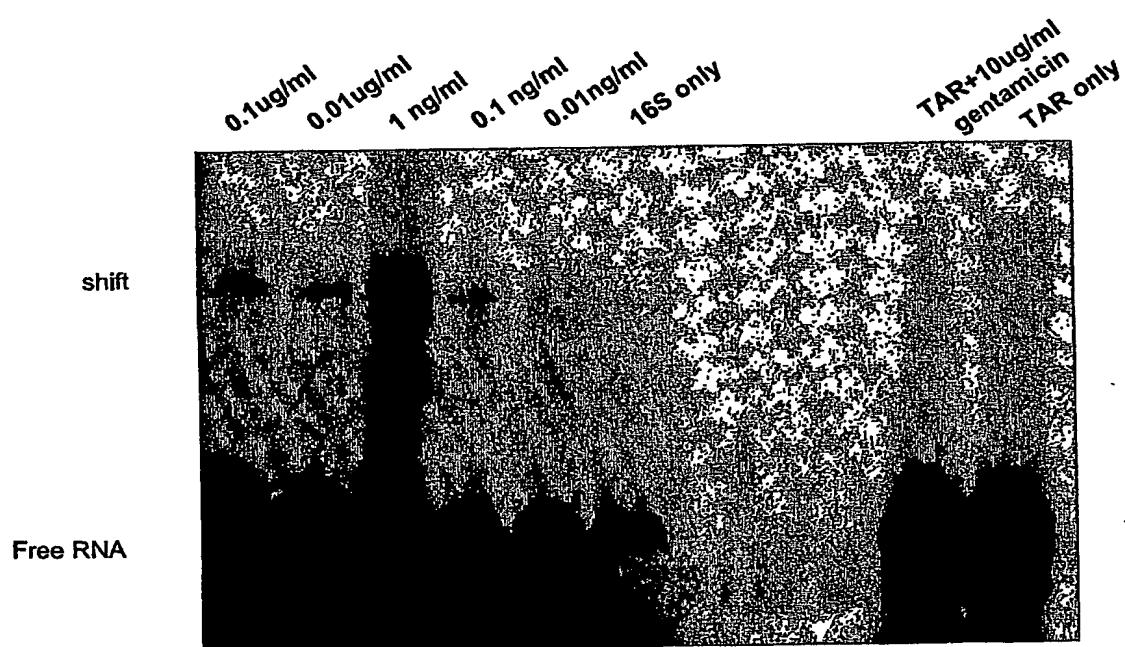


**Figure 3**  
**Sheet 3/11**  
**Attorney Docket No. 10589-021**



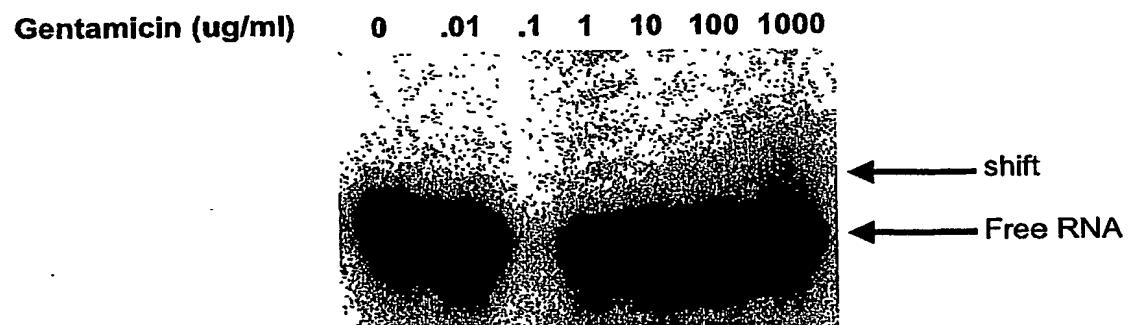
60398344.072402

**Figure 4**  
**Sheet 4/11**  
**Attorney Docket No. 10589-021**

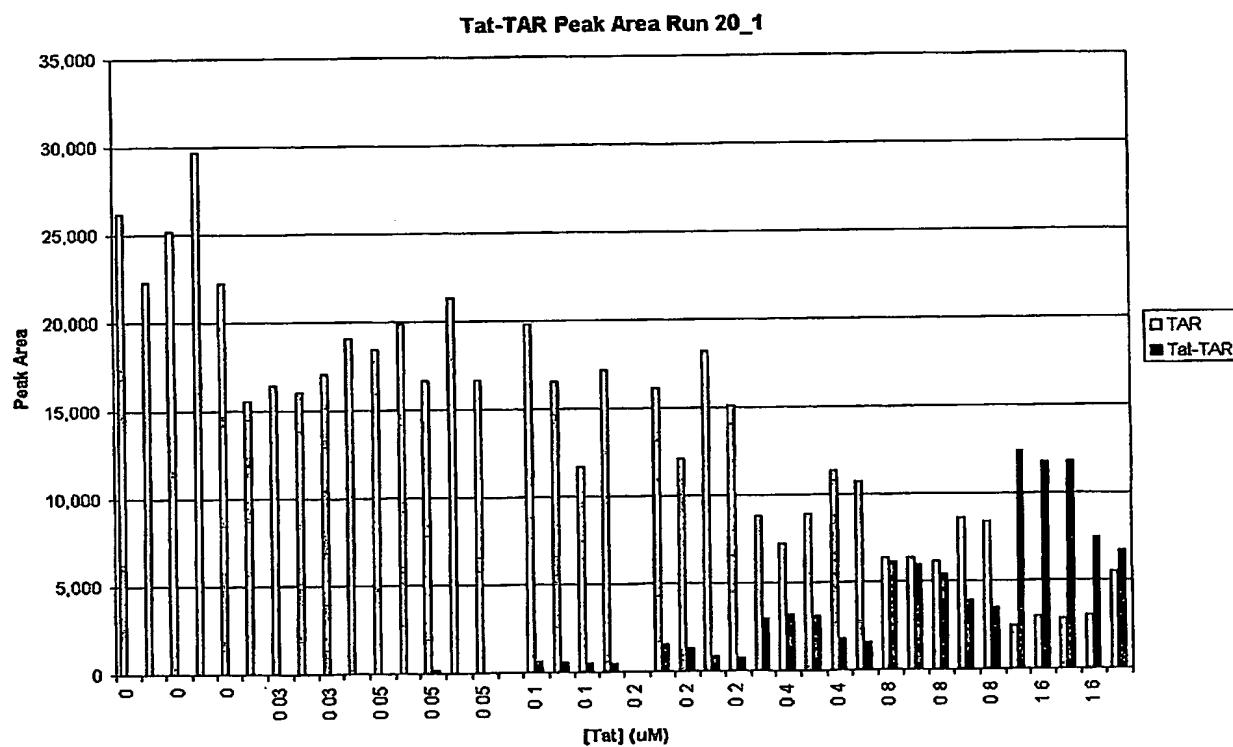


60396244.0724.02

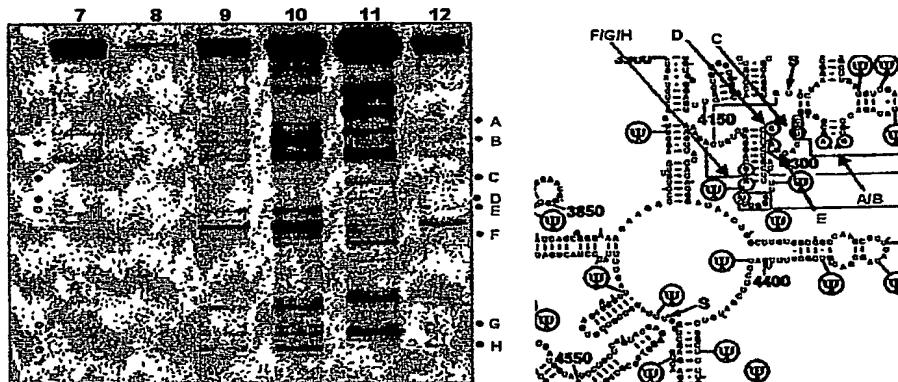
**Figure 5**  
**Sheet 5/11**  
**Attorney Docket No. 10589-021**



**Figure 6**  
**Sheet 6/11**  
**Attorney Docket No. 10589-021**



**Figure 7**  
**Sheet 7/11**  
**Attorney Docket No. 10589-021**



- 7. DMSO
- 8. Paromomycin
- 9. Compound A (A, B, E, F, G, H)
- 10. Compound B (A, B, E, F, G, H)
- 11. Compound C (C, D)
- 12. Compound D (A, B, E, F, G, H)

**Figure 8**  
**Sheet 8/11**  
**Attorney Docket No. 10589-021**

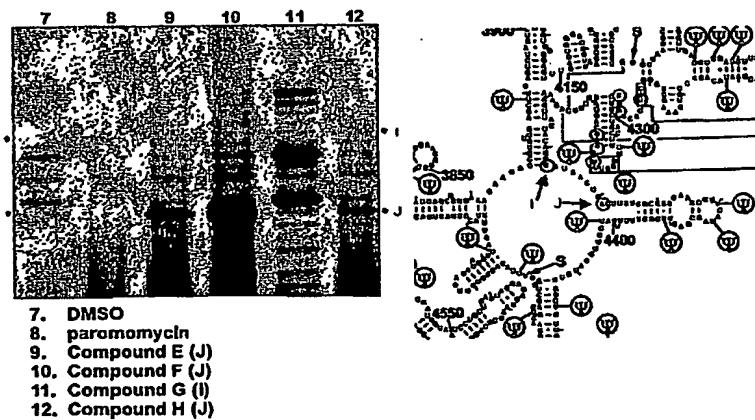
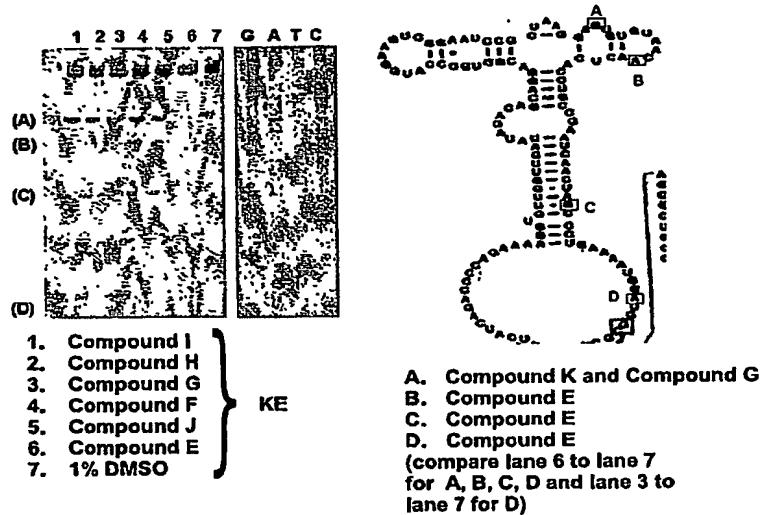
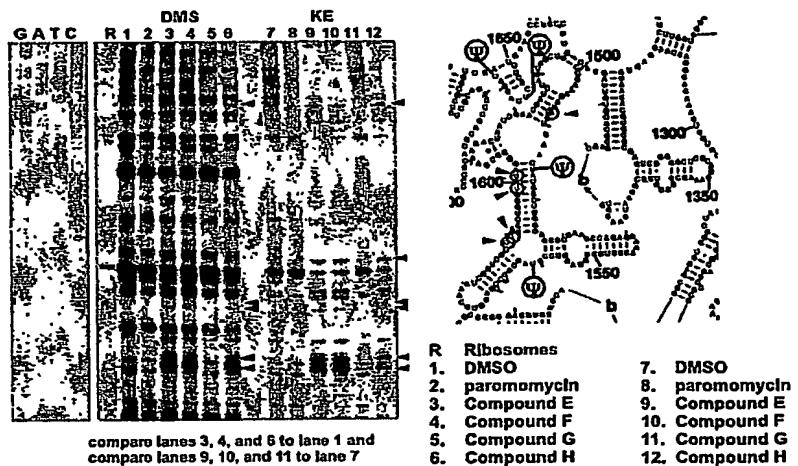


Figure 9  
Sheet 9/11  
Attorney Docket No. 10589-021



**Figure 10**  
**Sheet 10/11**  
**Attorney Docket No. 10589-021**



**Figure 11**  
**Sheet 11/11**  
**Attorney Docket No. 10589-021**

